For Reference

NOT TO BE TAKEN FROM THIS ROOM

For Reference

NOT TO BE TAKEN FROM THIS ROOM

Ex libris
universitatis
albertaensis

UNIVERSITY OF ALBERTA
QUAECUMQUE VERA

University of Alberta
Printing Department

Digitized by the Internet Archive
in 2018 with funding from
University of Alberta Libraries

https://archive.org/details/differentialeffe00brow

Thesis
1961(F)
#10

THE UNIVERSITY OF ALBERTA

THE DIFFERENTIAL EFFECT OF STRESS-INDUCING SUPERVISION ON CLASSROOM TEACHING BEHAVIOR

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DIVISION OF EDUCATIONAL ADMINISTRATION

ALAN FERGUS BROWN
CALGARY, ALBERTA
OCTOBER, 1961

ABSTRACT

The two-dimensional theory of interpersonal stress dynamics in educational supervision, developed and tested in this thesis, states that the optimum effectiveness of an individual working in an institutional environment occurs as a unique combination of (1) the pressures, demands and role-expectations of the institution as the external drive stimuli, with (2) relevant personality elements of the individual as the internal drive arousibility factors. The process of individual supervision, which attempts to create conditions conducive to the optimum effectiveness of the individual, is regarded as potentially stressful to the supervised and thus may intensify the external drive stimuli.

It was hypothesized that a combination of some quantifiable personality elements with a reasonably constant external stimulus situation would, when the external stimulus is intensified through supervisory methods that are stress-inducing to the teacher, result in quantifiable variations in the effectiveness of classroom teaching behaviour. Subjects whose internal drive arousibility factors (defined as level of neuroticism or anxiety, general intelligence or scholastic aptitude, and the need for achievement) were high, i.e. at or above the optimum, would exhibit a deterioration in teaching effectiveness while subjects who were low, i.e. below the optimum would improve. Those in the middle-range would deteriorate slightly.

The hypotheses were tested in a series of controlled experiments, one with each of 78 student teachers during their regular teaching activities in the public schools. The subjects were visited by experimenters who, under the guise of faculty consultants that supervise student teaching, observed the control lesson, verbally induced failure-stress, and indicated the degree and direction of change in effectiveness in subsequent teaching.

Statistical analysis showed that, under conditions of stress-inducing supervision as compared with control conditions,

1. there was a slight mean decrement in teaching effectiveness,

2. the teaching effectiveness of high-neuroticism subjects deteriorated the most, of medium-neuroticism subjects deteriorated slightly, and of low-neuroticism subjects improved significantly, and that the same was true of subjects classified by degree of anxiety,

3. the teaching effectiveness of high scholastic aptitude subjects deteriorated the most, of medium scholastic aptitude subjects deteriorated slightly, and of low scholastic aptitude subjects improved significantly.

4. change in teaching effectiveness was not related to strength of need for achievement nor to extraversion-introversion.

Multiple range comparisons of the effects of stressful supervision upon components of effective teaching showed that

(a) the more personal aspects of teaching deteriorated the most, (b) the more bookish aspects of teaching improved slightly, and (c) the more traditionally pedagogical aspects of teaching followed the general average of a slight deterioration.

The findings are construed as supporting the theory and emphasizing the need for administrative personnel to differentiate their supervisory procedures.

ACKNOWLEDGMENTS

The author wishes to express his sincere gratitude to a number of persons who have assisted in the preparation of this thesis. For their inspiration and encouragement as well as for their criticism and stabilizing influence, the author is greatly indebted to the members of this committee, Dr. J. H. M. Andrews, Dr. R. S. MacArthur, Dr. H. T. Sparby, and Miss L. N. Wilson. Without the help of fourteen graduate fellows of the Division of Educational Administration, the experiment could not have been conducted; their assistance is gratefully acknowledged. The author similarly acknowledges the receipt of a grant towards expenses of the experiment from the Research Council of the Canadian Education Association.

TABLE OF CONTENTS

Page

List of Tables . . . .

List of Figures . . . .

Chapter

I. INTRODUCTION 1

Stress: Desirable and Undesirable 2

Some Guiding Concepts (Assumptions) . . . 4

Specific Purpose of this Thesis 6

II. ADMINISTRATIVE RELATIONSHIPS IN A SYSTEM OF SOCIAL ACTION 12

Parsons' Action Theory 12

Getzels' Theory of Administrative Relationships 13

Sources of Conflict and Dissatisfaction . . 17

Contribution to a Solution of the Problem . 21

III. RESEARCH ON SOCIAL ACTION THEORY IN ADMINISTRATION 25

Three-Variable Studies 25

Partial Studies 31

IV. EXAMINATION AND EXTENSION 37

Stability and Fertility 37

Limitations in the Personal Dimension . . 39

Directional Nature of Predictions 42

V. THE PERSONAL DIMENSION 45

Carl Rogers and Administrative Relationships 45

The Variable of "Scholastic Aptitude" . . 51

Chapter Page

VI. DEFINING ANXIETY, NEUROTICISM, STRESS AND DRIVE. . 54
Usage in the Technical Literature 54
Classification of Stress Stimuli 59
Factors and Correlates of Anxiety and Stress. . . 61
Accepted Definitions of Terms Used in This Thesis 65

VII. THEORETICAL IMPLICATIONS OF STRESS-REACTIONS STUDIES 72
Drive Levels and Task Performance 72
Contribution to a Solution of the Problem 77

VIII. THE INTERPERSONAL DYNAMICS OF SUPERVISORY RELATIONSHIPS: A THEORETICAL ORIENTATION 83
Three Personality Factors 84
Neuroticism 84
Scholastic Aptitude or General Intelligence . . . 85
Need for Achievement 86
Directions of Performance Change Under Stress . . 89
Degree of Performance Change Under Stress 92
Mean Effect 94
Independent Variation 95
Implications for Supervision 95

IX. BASIC HYPOTHESES 100

X. OPERATIONAL DEFINITIONS: INSTRUMENTATION 105
Need for Achievement 106
Neuroticism and Anxiety 108
Scholastic Aptitude or General Intelligence . . . 113
Stress Situation 113
Performance Change 115

Chapter Page

Classification of Variables 119

XI. EXPERIMENTAL PROCEDURE: SUBJECTS, PROGRAM AND EXPERIMENTERS. 124

The Sample 125

Characteristics of the Sample 127

Collecting Test Data 129

The Student Teaching Program 129

Clearance and Conditions 131

The Experimenters 132

Scheduling 133

XII. PROCEDURE CONTINUED: THE EXPERIMENTAL VISIT . . . 136

General Routine 136

Instructions to Experimenters 141

Experimenter's Kit 142

Experimenter Training 148

Security 149

The Humaneness of the Project 152

XIII. DISTRIBUTION OF PERFORMANCE CHANGE UNDER STRESS. . 155

Confidence Coefficient = .95 155

Symbolization 156

Testing Hypothesis One. 158

Testing Hypothesis Two. 160

Testing Hypothesis Three 161

Chapter Page

XIV. SEPARATE EFFECTS OF PERSONAL FACTORS 165

Three Levels of Each Factor 165

Treatments Fixed, Not Random. 165

Simple Variance Design 167

Assumptions 168

Mathematical Model and Hypothesis 169

Testing Hypothesis Four 170

Testing Hypothesis Five 174

Testing Hypothesis Six 178

Need for Achievement 179

Implications of $F < 1$ 181

Achievement via Conformity 184

Achievement via Independence 186

XV. NEUROTICISM AND SCHOLASTIC APTITUDE: INTERACTION AND MAIN EFFECTS 191

Mathematical Model and Hypotheses 191

Obtained Two-Way Layout and Assumptions 193

Testing Hypotheses Four, Five and Seven 200

Effect of Non-Normal Cell_{31} 202

Effect of Unequal Cell Entries 204

Anxiety and Neuroticism 205

Extraversion-Introversion 207

Conclusions 209

Chapter Page

XVI. TREND ANALYSIS OF THE EFFECTS OF COMBINED DRIVE AROUSIBILITY INTENSITIES 212

Combination of Main Effects 212

Five Levels of Intensity 214

Analysis of Diagonals 215

Significance of Linearity Component 220

Conclusions 223

XVII. THE COMPONENTS OF TEACHING EFFECTIVENESS 227

Personal Qualities 228

Preparation for Teaching 230

Presentation of Lessons 233

Contact with Pupils 235

Classroom Management 239

Analysis of Variance between Component Means . . 243

Multiple Range Comparisons of Component Mean Differences 246

XVIII. SUMMARY AND CONCLUSIONS 253

Theory of Stress-Inducing Supervision Restated . . 253

Hypothetical Solutions 256

1. General Effect 256

2. Effect of Personality Differences 257

3. Effect upon Components of Teaching 258

General Results 260

General Conclusions 260

Summary of Specific Findings 262

Chapter Page

XIX. IMPLICATIONS 267

Implications for Improved Educational Practice. . 268

Theoretical Implications 273

Implications for Future Research 276

BIBLIOGRAPHY 281

APPENDIX: TESTS USED IN THE INVESTIGATION 296

1. Maudsley Personality Inventory (MPI-N and MPI-E) 297

2. Students' Biographical Index (MAS, CPI-Ac, CPI-Ai, and EPPS-nAch) 300

3. Confidential Report on Student Teaching (Research Adaptation of Forms 60-2, 50-2) 307

LIST OF TABLES

Table Page

1. Distribution of D-Scores Showing Frequencies and Statistics Calculated from the Distribution, N=78 159

2. Goodness of Fit of a Normal Distribution with an Obtained Distribution of D-Scores Showing Teachers' Reactions under Stress, N=78 162

3. Distributions of D-Scores of Three Subpopulations Classified by Level of Neuroticism (MPI-N), N=78 . 172

4. Analysis of Variance of D-Scores for Three Subpopulations Classified by Level of Neuroticism (MPI-N) 173

5. Distributions of D-Scores of Three Subpopulations Classified by Level of Scholastic Aptitude (ACE), N=78 175

6. Goodness of Fit of Normal Distribution with an Obtained Distribution of D-Scores from 26 Subjects Classified as Medium-Level ACE 176

7. Analysis of Variance of D-Scores for Three Subpopulations Classified by Level of Scholastic Aptitude (ACE) 177

8. Distributions of D-Scores of Three Subpopulations Classified by Level of Need for Achievement (n. Ach), N=78 180

9. Analysis of Variance of D-Scores for Three Subpopulations Classified by Level of Need for Achievement (n. Ach) 181

10. Distribution of D-Scores of Three Subpopulations Classified by Level of Need for Achievement via Conformity (Ac) 185

11. Analysis of Variance of D-Scores of Three Subpopulations Classified by Level of Need for Achievement via Conformity (Ac) 186

12. Distributions of D-Scores of Three Subpopulations Classified by Level of Need: Achievement via Independence (Ai), N=78 187

Table Page

13. Analysis of Variance of D-Scores of Three Subpopulations Classified by Level of Need: Achievement via Independence (Ai) 188

14. Population Values Estimated by Mean Squares in an Experiment with r Rows, c Columns and n Individuals in Each Cell 192

15. D-Scores of 78 Student Teachers Submitted to Stress-Inducing Supervision, Classified by Level of Neuroticism (MPI-N) and by Level of Scholastic Aptitude. (ACE). 194

16. Sums of Squares and Variances of Nine Subpopulations Classified by MPI-N and ACE 195

17. Goodness of Fit of Normal Distribution with Each of Nine Subclass Distributions Classified by MPI-N and ACE 197

18. Analysis of Variance of D-Scores Shown in Table 15 201

19. Goodness of Fit of Hypothetical Distribution$_{31}$ with a Normal Distribution, $N_{31} = 8$ 202

20. Analysis of Variance of Hypothetical Distribution of D-Scores (One Score Changed from Table 15) 203

21. Analysis of Variance of D-Scores Shown in Table 15, Calculated by the Method for Unequal Frequencies in Subclasses 205

22. Two-Way Classification of Mean D-Scores Showing Five Levels of Intensity of Treatment-Combinations. . . 214

23. Mean D-Scores of 78 Student Teachers Classified by Intensity of Drive Arousibility Factors (MPI-N, ACE) 216

24. Trend Analysis of Diagonal Cells from Table 15 . . 219

25. Orthogonal Coefficients and Other Data Necessary for Testing the Significance of Linearity between Five Subclass Means 222

26. Test for Significance of Linear Regression and Deviations from Linear Regressions for the Data of Table 25 224

27. Sums, Frequencies and Means of 78 D-Scores for Personal Qualities, Classified by Level of Neuroticism (MPI-N) and Scholastic Aptitude (ACE) 229

Table Page

28. Analysis of Variance of D-Scores for Personal Qualities from Data in Table 27 230

29. Sums, Frequencies and Means of 78 D-Scores for Evidence of Preparation for Teaching, Classified by Level of MPI-N and ACE 231

30. Analysis of Variance of D-Scores for Preparation for Teaching from Data in Table 29 232

31. Sums, Frequencies and Means of 78 D-Scores for Presentation of Lesson, Classified by Level of MPI-N and ACE 234

32. Analysis of Variance of 78 D-Scores for Lesson Presentation from Data in Table 31 235

33. Sums, Frequencies and Means of 78 D-Scores for Contact with Pupils, Classified by Level of MPI-N and ACE 236

34. Analysis of Variance for Contact with Pupils, from Data in Table 33 238

35. Sums, Frequencies and Means of D-Scores for Classroom Management, Classified by Level of MPI-N and ACE 240

36. Analysis of Variance for Classroom Management, from Data in Table 35 241

37. Data from Five Criterion Measures of Aspects of Teaching Effectiveness, Applied to Each of 78 Student Teachers in a Failure Stress Experiment . . 244

38. Analysis of Variance of Data from Table 37 245

39. Table of Significant Studentized Ranges (.01 Level) and Shortest Significant Ranges for k=5 Means . . 248

40. Duncan's New Multiple Range Test Applied to the Differences between k=5 Treatment Means, from Data in Table 37 249

LIST OF FIGURES

Figure Page

1. Conceptualization of Administrative Relationships in a Social System 16

2. Four Possible Dimensions Mediating the Effect of the Social System upon Observed Behavior 17

3. Theoretical Model of Administrative Relationships 18

4. The Curvilinear Relationship of Drive Level to Effective Performance on Complex Tasks 90

5. Degree and Direction of Change in Performance Under Stress, Expressed as a Trend in Means of Subjects Classified by Drive Intensity 218

6. Significant and Nonsignificant Distances between Means of the Components of Teaching Effectiveness 250

CHAPTER I

INTRODUCTION

The purpose of this study is to learn the effects upon teaching performance of certain administrative expectations expressed in a way that induces interpersonal stress, when individual differences among teachers in terms of personality needs and drives are considered.

From time immemorial, students have talked about "the neurotic teacher" and teachers have complained of job pressures and tensions. Administrators, taking an interest, ponder whether psychological stress and neurotic tendencies -- apart from being a nuissance to the smooth operation of the school -- really make much difference to the actual job of teaching. Their interest is well-taken, for it would seem that supervisory personnel need to know just as much about individual differences in teachers in order to provide differentiated supervision as teachers do about pupils to provide differentiated instruction. Unfortunately, there is very little, if anything, written about the differential effects of supervision resulting from personality and ability differences among teachers, although there does appear to be an increasing awareness that the processes of supervising and of educating hold much in common.

Stress: Desirable and Undesirable

Psychological stress, or neurotic anxiety, or emotional tensions* are often said to be growing phenomena of present day living. The pressures of modern organization (ref. 23), the concern that our society should be at once both "sane" and "affluent" (ref. 10, 11), the age of nuclear destruction and space conquest, all have at one time or another been said to account for the increase in anxiety tensions. Preoccupation with anxiety led a popular newsmagazine to comment:

> Anxiety seems to be the dominant fact -- and is threatening to become the dominant cliché -- of modern life. (3, p. 46)

On the other hand, anxiety is hardly new. Burton's Anatomy of Melancholy, which gives excellent descriptions of the condition, was a best-seller in the 1600's. Montaigne treated it casually as though it were a common thing in his culture. Shakespeare's plays include many examples of stress phenomena. Jersild sums it up:

> Ours is called an age of anxiety, and so it is. Each age is an age of anxiety. Ever since man has been affected by the stress of conflicting tendencies within himself, he has been anxious. (13, p. 20)

Indeed, although anxiety and tension are frequently thought of as undesirable states, this is not invariably the case. Most persons have, through experience and training,

---

*In this section the terms are used loosely. Indeed, even within the technical literature there sometimes appears a confusion of meanings. For formal definitions, see Chapter Six.

learned to accept them and even to consider them pleasurable under certain circumstances. For example, a race track habitué deliberately invites tension reactions by betting on the horses. At times, the absense of such reactions is unpleasurable. A psychiatrist recounted the following:

> One of our patients, an enthusiastic trout fisherman trained himself to relax when he had a fish on the line and reported that it spoiled all the fun. (12, p. 2)

In education, too, the necessity of a certain amount of anxiety tension is recognized. Miller makes the point that "Completely satisfied people are poor learners." (15, p. 21) A state of apathetic indifference carries weak motivation. On this basis, teachers have sometimes created in their students a state of fear and restlessness, martinette administrators have filled teachers with uncertainty and dread expecting, it would seem, that motivation would be sufficient to produce the desired results. Fear motivates, that is true; but it may motivate one to do something entirely unpredicted (see 4, p. 138). A certain level of stress is an energizer (14), but most educators would take the view that the challenge of a new learning situation and the intricacies of the teaching task themselves carry sufficient stress so that supplement by threat is less than necessary (20). In fact, it has been demonstrated that severe levels of anxiety result in a distinct decrement in complex task performance (18).

It is this possibility of a decrement in task performance that prompted the present study. While supervisory

personnel probably do not carry out individual supervision with the expressed purpose of arousing severe anxieties in the teacher, there is always the possibility of some emotional tensions being present. Whether this results in a performance deficit or, perhaps because of increased stimulation, an increment in effectiveness is a question in which administrators are interested. The investigation reported in this thesis attempts to answer just such a question.

Some Guiding Concepts (Assumptions)

The following four concepts are the assumptions on which the development of this thesis rests; additional assumptions will be introduced at the appropriate point.

1. Teachers are humans; thus there are individual differences in personality characteristics among them. In any group of individuals there will be wide variations in the degree to which they exhibit neuroticism or anxiety, in the amount of motivation they bring to their work, in their peculiar patterning of attitudes, in fundamental needs and in the self-images which they seek to maintain and enhance in their daily acitivity.

2. Teachers experience certain tensions, dissatisfactions and fears in the teaching job, and these become part of the overall stress of teaching. Ahrens (1), for example, found seven fears to be prominent among the teachers he studies:

i. fear of being unable to do an acceptable job,

ii. fear of criticism from parents and others of the community,

iii. fear of experimenting,

iv. fear of not being liked by the children,

v. fear of comparison with other teachers and of lack of acceptance by them,

vi. fear of administrative authority, and

vii. fear of losing their jobs.

Stevenson and Milt (19) discovered ten sources of teacher tensions: feeling inadequate as a teacher, petty frustration-producing annoyances, burying worry, worrying about oneself, becoming upset over failure of plans, feeling slighted or left out, lack of aggressiveness when criticized, overzealous ambition, and overwork. Aikenhead (2) found the following to be sources of discouragement among beginning teachers in Western Canada: discipline problems, critical parents, inadequate school facilities, inability to apply professional skills, dull routine, isolation, and peer ostracism. Such lists as these indicate the presence of stress-producing factors both situational and persistent within the individual's personality, within the administrative system in which he works, and within his culture.

3. Teaching may be viewed as the utilization of personal attributes in complex task performance. Inasmuch as successful teaching requires forming continuous solutions to successive, interrelated challenges or problems, it may be considered to be a complex, cognitive task. Similarly, to teach children

requires drawing on one's personality resources (the term, as used here, denotes both the cognitive and conative aspects of personality). It may, of course, be argued that much of teaching is routine rather than complex and interpersonal; the routine, however, is not the essence of good teaching. An advantage of attending to these two aspects of teaching behavior is that doing so immediately invokes two large bodies of literature which may provide concepts and techniques to assist in solving the present problem. The two, respectively, are theories of learning and personality psychology. Employing personality concepts to gain insight into problems of learning, or of task performance, under stress conditions has been characteristic of several recent writers, notably Dollard and Miller (5), Mowrer (16), Spence and associates (17, 18), and Eysenck (6, 7, 8, 9).

4. Individual supervision (here distinguished from group or cooperative supervision is potentially stressful to teachers. Any interpersonal event wherein there is a perceived superior-subordinate relationship, and where the focus of the event is the subordinate's activity relative to the superior's expectations, is assumed to contain interpersonal stress, at least on the part of the subordinate. The intensity of the stress may be so weak as to be negligible, or so severe as to affect behavior, depending upon a number of factors, several of which will be described later in this thesis.

Specific Purpose of this Thesis

The assumptions of the preceding section permit a

theoretical basis for the notion that several factors contribute to the drive level of a teacher in any given situation. Some of them are within the teacher: the teacher's desire to do a good job, her anxious or neurotic tendencies, and her self-expectations, based on previous experience, of being able to do a good job of this sort. Some of them are within the situation: the normal challenge of the job, the sources of tension, fear and discouragement that are commonly found in teaching. Adding individual supervision increases the overall drive level of the teacher at that time. The purpose of this thesis to determine the nature and extent of the effect upon the teacher's classroom behavior of a marked increase in interpersonal stress.

The factors that are "within the situation" may be relatively constant from one teacher to another. But since there is no reason for supposing the same with respect to those factors that are "within the teacher", it is quite possible that no two teachers will react in the same way when confronted with stress-provoking methods of supervision, that is where there is a marked increment in the stress factors 'within the situation.' The teaching of some may deteriorate, of some others it may actually improve, while still others may remain seemingly unaffected. This is because, in the absence of any knowledge of those factors 'within the teacher,' one must assume that to some persons the stress is an energizer but to others it is personally catastrophic. What would happen under repeated, long-term exposure to the kind of superior-subordinate conflict that is stressful is another question and one the

answer to which is not attempted in the present investigation.

An interesting inference of the foregoing statement could be that any event, activity or program which raises the drive level of the teacher may produce the same results, varying with the intensity of the event. If that were true, the psychological dynamics of in-service education activities could be revealed. Some direction regarding the optimum extent of participation sought and the methods of conducting the activity could be provided. There are, after all, ways of increasing psychological drive other than a direct "bawling out."

The controlled experiment forming the basis of this study in administrative relationships was designed to reveal, in generalizable terms, what happens to classroom teaching under conditions of interpersonal stress, experimentally induced to simulate that which may occur during the process of individual supervision, and to analyze these effects according to the individual differences found among the subjects. For example, the question would be asked: Do subjects classed as very neurotic exhibit a teaching reaction that differs from those classed as very stable, or from those classed as average in their neurotic tendencies? Or, again, compared with "normal teaching conditions, does the necessaty of teaching under duress have an effect upon the performance of persons classed as having high scholastic aptitude (and, inferentially, past history that would build a self-image of success on cognitive tasks) that is distinct from the performance of subjects classed at the lower end of the scale, or the middle? Or, more ambitiously:

If there is a difference, how does one account for it?

Although the specific purpose is contained in the preceding paragraph, the study is also expected to shed light on questions commonly asked in two larger areas of investigation: (1) the relation between employee performance and the congruency between (a) his personal needs or attitudes and (b) the institutional expectations or standards held for his role; and (2) the effects of anxiety and stress upon learning in particular and upon complex in more detail the backgrounds in each of these two fields, ending with an attempt to integrate their basic concepts into a unifying two-dimensional theory of interpersonal stress dynamics. Following these developmental chapters will be found the report of an experiment conducted to test certain hypotheses deduced from the theoretical prediction and explanation of teacher behavior under stress.

REFERENCES FOR CHAPTER I

(1) Ahrens, M. R. "Helping Teachers with their Tensions," Childhood Education, 28 (Jan. 1952) 211-215

(2) Aikenhead, J. D. "Teacher Satisfactions and Discouragements," Alberta Journal of Educational Research, 4:2 (June 1960) 92-102

(3) "Anatomy of Angst, The," Time Magazine, 77:14(March 31st, 1961) 46-51.

(4) Clarke, S. C. T. "School Control and Discipline," The 1956 Alberta School Principal, Division of Educational Administration, University of Alberta, 1956, 134-145.

(5) Dollard, J. and N. E. Miller. Personality and Psychotherapy. New York; McGraw-Hill, 1950.

(6) Eysenck, H. J. The Structure of Human Personality. London: Methuen, 1953.

(7) Eysenck, H. J. "A Dynamic Theory of Anxiety and Hysteria," Journal of Mental Science, 101:(Jan. 1955) 28-51.

(8) Eysenck, H. J. "Reminiscence, Drive, and Personality Theory," Journal of Abnormal and Social Psychology, 53 (1956) 328-333.

(9) Eysenck, H. J. The Dynamics of Anxiety and Hysteria. London: Routledge and Kegan Paul, 1957.

(10) Fromm, Erich. The Sane Society. New York: Rinehart, 1955.

(11) Galbraith, John K. The Affluent Society. London: Hamish, 1958.

(12) Haugen, G. B., H. H. Dixon and H. A. Dickel. A Therapy for Anxiety Tension Reactions. New York: MacMillan, 1958.

(13) Jersild, Arthur T. When Teachers Face Themselves. New York: Bureau of Publications, Teachers College, Columbia University, 1955.

(14) McKinney, F. "Emotional Factors in Learning," Journal of General Psychology, 9 (1933) 101-116.

(15) Miller, N. E. and J. Dollard. Social Learning and Imitation. New Haven: Yale University Press, 1941.

(16) Mowrer, O. H. Learning Theory and Behavior. New York: Wiley, 1960.

(17) Spence, K. W. and J. A. Taylor "The Relation of Conditioned Response Strength to Anxiety in Normal, Neurotic, and Psychotic Subjects," *Journal of Experimental Psychology*, 45 (1953) 265-272.

(18) Spence, K. W., J. A. Taylor and R. Ketchel "Anxiety (Drive) Level and Degree of Competition in Paired Associate Learning," *Journal of Experimental Psychology*, 52 (1956) 306-310.

(19) Stevenson, G. S. and H. Milt "Ten Tips to Reduce Teacher Tension," *National Educational Association Journal* 45 (Dec. 1956) 545-547.

(20) Symonds, P. M. "What Education Has to Learn from Psychology, VI: Emotion and Learning," *Teachers College Record* 60:1 (Oct. 1958) 9-22.

(21) Taylor, J.A. "Drive Theory and Manifest Anxiety," *Psychological Bulletin* 53 (1956) 303-320.

(22) Taylor, J. A. and K. W. Spence "The Relationship of Anxiety Level to Performance in Serial Learning," *Journal of Experimental Psychology*, 44 (1952) 61-64.

(23) Whyte, W. *The Organization Man*. New York: Simon and Schuster, 1956.

CHAPTER II

ADMINISTRATIVE RELATIONSHIPS IN A SYSTEM OF SOCIAL ACTION

The most comprehensive formulation of administrative relationships has been Getzels' psycho-sociological framework of organizational roles and individual needs (2, 3, 4, 5, 6, 7, 8). Its antecedents would appear to be the successful integration of Henry A. Murray's need-press theory in psychology (12) with Talcott Parson's theory of social action (14). Murray long ago established a firm basis for the notion that an individual's behavior in a social situation was a function of his unique structure of needs, a number of external "presses", and the interaction between the two. A press was a force originating from within. In our present problem in administrative relationships, Murray would have referred to external expectations as the "institutional press."

Parsons' Action Theory

In Parsons' theory of social action the bejavior of an individual in a social situation is determined by (1) his conformity to his own standards (or needs) and (2) his conformity to social standards (presses). Such social standards in administrative relationships are, chiefly, those of the formal organization: the administrative expectations held for the individual's role in that organization. The theory is described by Parsons as being "structural-functional," that is,

> . . . it includes a system of structural categories which must be logically adequate to give a determinate description of an empirically possible, complete empirical system of the relevant class. . . . On the other hand, such a system must include a set of dynamic functional categories. . . . (13, p. 23)

As to content, Selznick says the theory may be characterized in a single sentence, "Structural-functional analysis relates contemporary and variable behavior to a presumptively stable system of needs and mechanisms." (15, p. 29) In a problem in administrative relationships this would mean that the school system's goals (needs) and formal structure must be studied along with the personality dynamics of the individual teachers in that school system. In Toward a General Theory of Action, Parsons draws attention to the fact that action theory encompasses, and thereby integrates, all the main elements of sociological theory, personality psychology and social psychology.

Getzels' Theory of Administrative Relationships

Getzels' psycho-sociological framework has applied Parson's theory to administrative processes in the school, and specifically to the process of mediating between institutional expectations and personality needs to achieve the goals of the school. Accordingly, it offers hope for a solution to the problem of this thesis when stated this way: Against the standards of the school's expectations, what will be the individual teacher's observed behavior under conditions where there is a perceived conflict between role-expectations and individual needs or tendencies?

The basic notion of Getzels is that people who are brought together under some common purpose tend to develop some kind of a social system. Analyzing this social system reveals two prominent characteristics: (1) there is an institution with a pattern of roles among its members as well as a definite set of goals which the institution wants to accomplish; these goals become articulated for the individual members in the form of role-expectations held for their behavior, (2) there are the individual, flesh-and-blood persons to whom are assigned these roles; these persons are part of the system but are, nevertheless, individuals with their own idiosyncratic personalities and need-dispositions. These two characteristics although they may be separated for conceptual purposes, are interrelated phenomenologically. The first characteristic could be referred to as the "normative dimension"; the second the "personal dimension."[1] Behavior, therefore, is a function of the interaction between the normative and personal kinds of events taking place within a social system. Going a step further, it is possible to analyze each of these two dimensions in terms of their analytic units. Thus an institution is a system of roles which is a system of role-expectations. Similarly, the individual is (so far as this analysis requires) his personality which is made up (according to Getzels) of a

---

[1] The corresponding terms "nomothetic" and "idiographic" are customarily employed throughout Getzels and many others. This labouring with words adds little.

system of need-dispositions. "In a sense," Getzels has remarked "the one may be thought of as the 'sociological' level of analysis the other the 'psychological' level of analysis." (8, p. 66).

Because there is often an ambiguity in the terms "personality" and "need-disposition", it is noted that, for Getzels, personality is

> the dynamic organization within the individual of those need-dispositions that govern his unique reactions to the environment and . . .to the expectations in the environment. (8, p. 68)

He adopts the definition of a need-disposition provided by Parsons and Shils as

> individual tendencies to orient and act with respect to objects in certain manners and to expect certain consequences from these actions. (14, p. 114)

In summary, Getzels' theory of administrative relationships may be presented in three ways: statement, equation and diagram. His statement:

> . . . social behavior results as the individual attempts to cope with an environment composed of patterns of expectations in ways consistent with his own independent pattern of needs.
> (8, p. 69, also 3, p. 157)

The equation:

$$B = f(RxP)$$

where:

B is observed behavior,

R is a given institutional role defined by the expectations attaching to it, and

P is the personality of the particular role-
incumbent defined by his need-dispositions

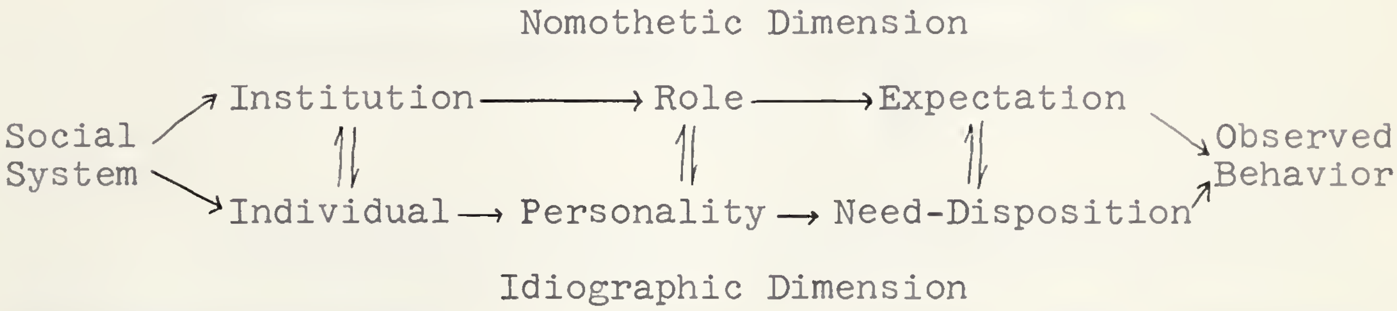
(3, p. 157, also 8, p. 69)

The diagram is shown in Figure 1 (2, 3, 4, 8).

Figure 1

A Conceptualization of Administrative

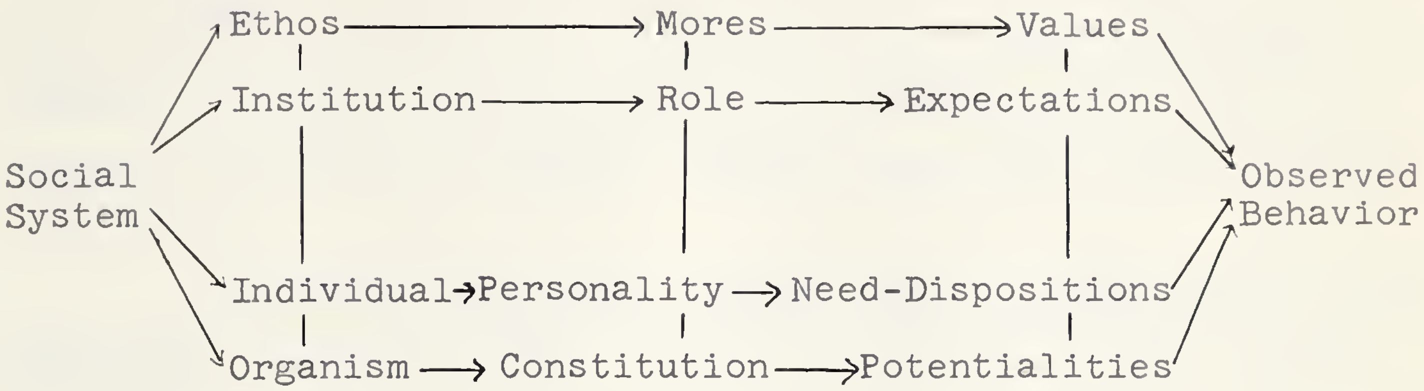
Relationships in a Social System



A general theory should, to be sure, embrace all possible relevant variables. Those presented in the diagram are fairly exhaustive but Getzels hastens to note that behavior is determined also by variables in two additional dimensions: those suggested respectively by the fields of biology and anthropology. Although the cumbersome term "psycho-socio-bio-anthropological framework" has not been applied to pictorial representation of Figure 2, it would serve to emphasize that any study of individual reactions to administration-induced tensions must be interdisciplinary in character.

Figure 2

Four Possible Dimensions Mediating the Effect of the Social System upon Observed Behavior



(8, p. 73)

The model(s) achieves the desired quality of being structural-functional. It takes into account both the pattern of interactions among the major elements of the situation, and the interpersonal dynamics of the real persons making up the system. But how does it work?

Sources of Conflict and Dissatisfaction

When (a) the role-expectation held for the individual by the organization (or its agent, e.g. supervisor), as perceived by the individual (whether correctly or incorrectly) is in any way in conflict with (b) the individual's need-dispositions, then "expectations" and "needs" are said to be incongruent. When there is agreement, they are congruent. Ideally, the social system would be such that the individual fulfils both the expectations held for him by the institution and the needs of his own personality. In this regard, Hills makes the practical point that:

> Obviously this can only be the case when the expectations and needs are perfectly congruent -- a situation rarely, if ever, found in practice. There is inevitably a greater or lesser amount of strain or conflict between the individual and the demands of the institution. (10, p. 2)

Hills' comment, of course, agrees with the suggestion made in Chapter One that a certain amount of conflict and strain (i.e. stress) is normal in teachers working within any school system.

Getzels' theory and diagram specify three means through which this conflict arises. This is best exemplified, however, in an adaptation presented by Guba and Bidwell (9, p. 5) that permits a closer examination of the dimensions contained in the general statement.

Figure 3

Theoretical Model of Administrative Relationships

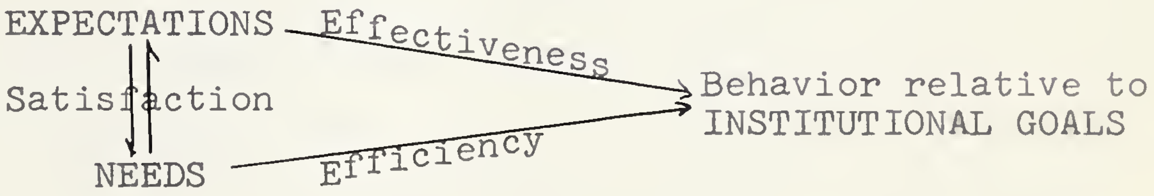


Figure 3 illustrates three areas of conflict: between those factors defining the individual's effectiveness, between those factors defining the individual's efficiency, and between those factors defining the individual's satisfaction in the job. The degree of congruence obtaining between the two factors at either end of a dimension determines the amount of conflict or lack of it, as well as the level of effectiveness, efficiency or satisfaction at which the individual will operate. These

terms may be defined as follows:

Effectiveness is simply the extent to which the individual attains the goals set for him by the institution. The model, however, recognized that behavior ratings are relative to the expectations held by the rater; this is in accordance with the view of personality assessment advanced by Stern, Stein and Bloom (16).

Efficiency is the degree to which the individual finds need-fulfilment in the process of carrying out the goals of the institution. Thus, behavior is efficient to the extent to which it is congruent with personality structure or needs. Similarly, inefficiency refers to the strain that results from an individual being compelled to behave in a manner that conflicts with his personal needs and energy that should go toward fulfilment of the task must be expended to maintain personality structure, or to reduce tension. This explication of the terms effectiveness and efficiency, while not always in accord with common usage which often makes them synonymous, was advanced by Barnard (1) and has enjoyed wide acceptance in the technical literature since.

Satisfaction may be defined as the relationship between expectations and needs. When the institutional expectations and personality needs are not congruent, satisfaction cannot be maximized. For example, the teacher with a high need for autonomy is not likely to be satisfied under close supervision. If we suppose that a teacher perceives that his principal or

supervisor expects him to carry out an assignment in a manner that conflicts with strong desires unique to this teacher, we should expect him to become dissatisfied and either (a) perform to the organization's dissatisfaction (i.e. ineffectively) but at no sacrifice to his self-perceptions (efficiently), or (b) perform to the organization's satisfaction (effectively) but with a great deal of strain to his self-perceptions (inefficiently). Since these two choices rarely, if ever, occur in pure form, there is normally a combination of some degree of both ineffectiveness and inefficiency present. Hence, one may draw the hypothesis that level of teacher performance relative to institutional goals is a function of the congruence between expectations and needs.[1]

This is, of course, "cold prediction:" Hypothesizing on the basis of elements as they are observed and assuming the emergence of no interfering or remedial factors. "Warm prediction", on the other hand, may indicate two further possibilities (note: for a discussion of "warm" and "cold" prediction, see 11, p. 67). The independent incongruent variables may alter in the direction of greater harmony -- such is one important task of supervision. Furthermore, our hypothetical teacher, even without supervisory assistance, may alter his self-perceptions in a direction resulting in a greater

---

[1] this is not an hypothesis of the present thesis; rather it is one that is logically deduced from the theory presented (and which, it will be seen shortly, has been deduced and tested elsewhere) in order to suggest one possible answer to the questions raised in Chapter One.

congruence with expectations. This "do-it-yourself" personality change may simply result from the presence of other stronger and more stable needs such as the need to hold a job, the desire for promotion or professional acceptance, and so forth. It may also result, it is important for the supervisor to note, from a change in the teacher's perceptions of expectations arising from greater familiarity with them or with the supervisors representing them, thus clearing erroneous perceptions which were felt as conflicting.

Contribution to a Solution of the Problem of this Thesis

Lest this extended exposition of the expectations-needs-goals theory be considered tangential, its application to the original problem of this thesis will be reviewed.

Teachers are interested in learning the role-expectations that the school holds for them. There are a variety of means that the administration has at its disposal for "role-education" the teachers in its charge because defining these expectations is a primary responsibility of the administration (shared, in some respects, with teacher-education institutions and society at large). Supervisory personnel (principals, supervisors, inspectors, superintendents or others) are the teacher's personal contact with the administration so that an important function of supervision is role-educating. Through the processes of consultative and motivational activities, the supervisor acquaints the teacher with the expectations held for his role and, deliberately or not, with the extent to which the teacher is meeting these requirements.

Supervisory activities are both individual and group in nature; the chief interest at the present is to focus on individual supervision. If, during or as a result of individual supervision, there develops a conflict between some aspect or aspects of the total personality of the teacher and the expectations as perceived by him, then one may at once pose the question of what this stress-inducing supervision does to the classroom teaching. But since it is the class situation described by the expectations-needs-goals theory one may conclude that, relative to institutional goals, this teacher's classroom behavior will deteriorate.

This would appear to be a rather reasonable hypothesis and, considering the massive theoretical basis on which it is placed, should answer the basic question of this thesis. First, however, two conditions must be met: (1) as an hypothesis it requires experimental verification, and (2) as an hypothesis it requires critical examination to reveal possible inadequacies in its application or derivation.

Should sufficient experimental verification be found, the first condition will be satisfied. The fact is, however, that despite its widespread acceptance and rather general application, experimental evidence in support of this theory is unimpressive.

REFERENCES FOR CHAPTER II

(1) Barnard, Chester. The Functions of the Executive. Cambridge: Harvard University Press, 1938.

(2) Getzels, J. W. "A Psycho-sociological Framework for the Study of Administration," Harvard Educational Review 22 (Fall 1952) 235-246.

(3) Getzels, J. W. "Administration as a Social Process," Administrative Theory in Education, A. W. Halpin, ed., Midwest Administration Centre, 1958.

(4) Getzels, J. W. Theory and Practice in Educational Administration: An Old Question Revisited," Administrative Theory as a Guide to Action, R. F. Campbell and J. M. Lipham, Eds. Chicago: Midwest Administration Centre, 1960, pp. 37-58.

(5) Getzels, J. W. and E. G. Guba "Role, Role Conflict and Effectiveness: An Empirical Study," American Sociological Review, 19 (1954) 164-175.

(6) Getzels, J. W. and E. G. Guba "The Structure of Roles and Role Conflict in the Teaching Situation," Journal of Educational Sociology 29 (Sept. 1955) 30-40.

(7) Getzels, J. W. and E. G. Guba "Social Behavior and the Administrative Process," School Review, 64:5 (Winter 1957) 423-441.

(8) Getzels, J. W. and H. A. Thelen "The Classroom as a Unique Social System," The Dynamics of Instructional Groups: The 59th N.S.S.E. Yearbook, Part II. Chicago: The Society, 1960, 53-82.

(9) Guba, E. G. and C. E. Bidwell. Administrative Relationships: Teacher Effectiveness, Teacher Satisfaction and Administrative Behavior. Chicago: Midwest Administration Centre, 1957.

(10) Hills, R. J. "A New Concept of Staff Relations," Administrator's Notebook, 8:7 (March 1960) 1-4.

(11) MacArthur, R. S. "The Coloured Progressive Matrices as a Measure of General Intellectual Ability for Edmonton Grade III Boys," Alberta Journal of Educational Research, 6:2 (June 1960) 67-75.

(12) Murray, Henry A. Explorations in Personality. New York: Oxford University Press, 1938.

(13) Parsons, Talcott. Essays in Sociological Theory Pure and Applied. Glencoe, Ill.; The Free Press, 1949.

(14) Parsons, Talcott and E. Shils. Toward a General Theory of Action. Cambridge: Harvard University Press, 1951.

(15) Selznick, Phillip "Foundations of the Theory of Organization," American Sociological Review.

(16) Stern, G. G., M. I. Stein, and B. S. Bloom. Methods in Personality Assessment. Glencoe, Ill.: The Free Press, 1956.

CHAPTER III

RESEARCH ON SOCIAL ACTION THEORY IN ADMINISTRATION

If research is to support the equation $B = f(RxP)$, that effective behavior depends upon a needs-expectations congruency, one must adduce evidence that employs (a) a criterion of effective behavior, (b) clearly stated institutional expectations for the role of the teacher or, at least, the teacher's perceptions of them, and (c) data on the teacher's personality sufficient to deduce its degree of congruence with perceived expectations. Thus there are three classes of variables.

Three-Variable Studies

There are but three three studies that have employed all three variables and their interrelations; none is a controlled experiment and one is largely the replication of another. Yet this paucity of research is simply a reflection of the difficulties encountered in putting even the most promising notions in the complex field of educational administration to the operational test. Reviewing administrative theories generally, Griffiths recognized the difficulties and present status as follows:

> The more frequently observed inadequacy is the lack of operationism in the concepts being employed. . . . This is somewhat true even in the Getzels formulation which has the most operational definitions to date. . . . The relationship between observed behavior and expectations and need-dispositions is an inferential one which cannot easily be verified. (15, p. 68)

The three studies in point appeared in this order: Getzels and Guba (13) on the relationship between role conflict and effectiveness, Guba and Bidwell (16) on virtually all relationships expressed in the theoretical model, and Campbell (8) verifying much of Guba and Bidwell's work.

Getzels and Guba studied the "social system" of Air University, Maxwell Airforce Base, where there were possibilities of role conflict, personality conflict, and role-personality conflict, among the officer-instructors. They found that the extent of role conflict was reflected in the number and magnitude of incompatible expectations placed upon instructors. All courses were taught within the social system of a military base. Instructors, however, who taught civilian-type courses within this setting felt more role conflict than those who taught military-type courses there. Military course instructors were, of course, more military-oriented than were instructors of civilian courses in accounting or legalistics. Furthermore, there were greater conflict-scores for officer-instructors who did not volunteer for teaching duty, who were not permitted to refuse the assignment, and who had to teach subject-matter not in their interests or range of competence. This part of the study demonstrated that a teacher experiences conflict (its opposite was labelled 'satisfaction' in the model) when he perceives personal needs, wishes, interests and abilities to be transgressed by the institution, i.e. to be in conflict with role expectations held for him by the administration. Similarly the teacher experiences conflict when he perceives an incompati-

bility among the expectations held for him (in this case, between militaristic and professional orientations).

The criterion variable of officer-instructor effectiveness was then related to intensity of felt conflict. Statistical analyses indicated that the greater the officer-instructor's involvement in conflict, the greater was his relative ineffectiveness. The criterion measure for the ineffectiveness was a composite of the subjective ratings of one's fellow-instructors. Encouraged by the general results, the authors were led to conclude that:

> These data provide excellent supporting evidence for the validity of the theory, methods and instruments. (13, p. 172)

Guba and Bidwell reported on a study of the interrelatedness of the basic elements of the "theoretical model" which was carried out by the Midwest Administration Center at Chicago. This study included both a cross-sectional and a longitudinal (experimental) analysis.

The cross-sectional study used questionnaires devised to measure expectations, needs and goal-achievement, and their interrelations in terms of satisfaction and effectiveness. (It may be noted that one term--efficiency--is missing; this was necessary because of the difficulty of devising an adequate measuring instrument.) There were 24 schools and approximately 175 teachers surveyed. The hypotheses that they tested were all supported and subsequently rephrased as major conclusions which also reveal the purposes and procedures of the project. For

that reason, they are quoted verbatim:

> A generalized relationship exists between (a) the extent of the differences that appear in the definitions of the teacher's role as given by teachers and by principals and (b) the level of teacher effectiveness, satisfaction, and confidence in the administrative leadership. More specifically. . . . the following relationships seem to exist between teacher effectiveness and attitudes and administrative behavior:
>
> (a) Whatever the reality of the situation may be, the ratings of effectiveness which a principal gives to a teacher are a function of the degree to which he perceives that teacher conforming to the expectations which he holds for the teacher role.
>
> (b) Also independent of the nature of the "real" organizational world, a teacher's self-rating of effectiveness if a function of his perceptions of his principal's expectations as compared with his self-perceptions of job performance.
>
> (c) The satisfaction which a teacher expresses in his job is a function of the extent to which the institutional expectations perceived by the teacher are congruent with the expectations which the teacher feels ought to be held for him (idealized expectations).
>
> (d) The confidence in the principal's leadership which is exhibited by a teacher is a function of the congruence between the teacher's perceptions of administrative expectations and the teacher's idealized version of those expectations.
>
> (e) All aspects of staff relationships dealt with in this study -- satisfaction, confidence in leadership, and effectiveness on the job, -- seem closely related to the extent to which the perceptions, both of expectations and of behavior, held by principals and teachers coincide. (16, pp 65-68)

The longtitudinal study, during which some attempt was made to have the principals alter their administrative relationships in the direction of improved staff relations by increasing the degree of congruency among the basic variables over the course of the school year, yielded results which were

not at all clear-cut. This does not necessarily do damage to the basic theory, but rather it points to the difficulty in altering well-established personalities and interaction-patterns.

Campbell, again using questionnaire-survey instruments, studied expectations, needs, effectiveness, satisfaction and confidence in leadership. Again employing the congruence principle, he found that, in the schools he surveyed, the presence of a high degree of congruency of the principal's stated expectations for the role of the teacher and the teacher's wants in his own role was associated with (1) a high principal's rating of teacher effectiveness, (2) a high degree of job satisfaction stated by the teacher, and a high degree of teacher's stated confidence in the leadership of his principal. Thus, on the basis of the foregoing three studies, a measure of support has been established for Getzels' theory of administration as a social system.

It is at just this point that an awkward question may be put: What is the difference between congruency and conformity? From the Campbell study one may infer not only a potentially useful way of predicting effective behavior, but also a simple explanation of how apple-polishers gain merit-marks. Persons showing bovine contentment as well as confirmed "yes-men" could, it is predicted, demonstrate high confidence in leadership, high efficiency and high effectiveness but be chronic low-producers both now and in terms of organization growth. The difine discontent, the zealous rebel, and the enterprising

non-conformist will be constant deterrents to a loosely interpreted congruency formula.

On this same point, an interesting study would be one in which principal's ratings of teacher effectiveness were referred to the degree to which the teacher's personality obviously conformed to that of his principal, or did so far as the principal perceived. In a sense, this would test the hypothesis that a principal distributes effectiveness ratings according to the degree to which he sees teachers conforming to his own self-image. A recent study by Andrews and the present writer (3) tested the closely related hypothesis that a principal distributes his teacher-effectiveness ratings according to the degree of similarity between his scores on a battery of personality measures and those of his respective teachers. The weakness of this hypothesis is, however, that a principal's perception of a teacher's personality scores of which the principal is ignorant. The battery included the fifteen need-scores of the Edward Personal Preference Schedule, the six attitude scores of the Allport-Vernon-Lindsay Study of Values and the Minnesota Teacher Attitude Inventory. That the hypothesis received no support in the Andrews and Brown study may, therefore, be attributed either to the stated weakness or to the actual ability of principals to set aside their interpersonal similarities and differences when engaging in the professional task of teacher assessment. Earlier, Prince (19) produced somewhat equivalent findings: that teacher-principal similarities in value systems bore no relationship to either

the principal's ratings of teacher effectiveness nor to teacher satisfaction. This teacher-principal similarity did, however, bear a direct relationship to the teacher's confidence in the principal's leadership and to their ratings of his effectiveness.

Partial Studies

Reference is made here to some of the findings of studies which have explored certain specific relationships within the Getzels model. Their design, although less comprehensive that those of the previous studies. frequently permitted a closer analysis of one or two dimensions than may have been possible had all the dimensions been encompassed. Most of the findings were based on the questionnaire-survey method of data gathering; some included interviews or observation; none was a controlled-experiment. The following generalized results are construed as providing support for the Getzels conceptualization of administrative relationships:

1. When the teacher is faced with competing expectations held for his role, he experiences role-conflict and his job-satisfaction is relatively low (14, 5, 21).

2. When the teacher's perceptions of the expectations held for his role by the superior are incongruent with his expectations of leadership (i.e. need-dispositions relative to ideal leadership), he experiences role-personality conflict and his job-satisfaction is low (4).

3. When there is a lack of agreement among staff members in their attitudes towards leadership (i.e. conflicting need-dispositions within a staff), there is personality conflict and low job-satisfaction within a staff (18, 22, 23).

4. Satisfaction depends upon congruence or divergence of the teacher's expectations for the principal's role and his perceptions of the principal's actual behavior (4, 19).

5. Satisfaction depends upon convergence or divergence of the teacher's perceptions of the teacher's role, and the teacher's desires or wants for his own role (20, 4, 11).

6. The breaking-down of a teaching staff into two or more sub-cultures holding conflicting expectations for the teacher role tends to deter harmony and cause dissatisfactions (1, 2).

7. Role-conflict, or the competing expectations for the teacher-role held by different groups has a differential effect upon teachers relative to their personal and social characteristics (5, 22, 23).

8. Role-personality conflict is decreased and teacher-satisfaction increased when administrators share their functions, e.g. decision-making, with teachers, thereby altering role-perceptions and expectations in the direction of greater congruency (20, 9, 10).

In each of the above studies on authority relationship, usually principal-teacher, was present. Ferneau (12) studied consultants and administrators, the consultant bearing the same relationship to the administrator as the principal to the teacher but with the authority relationship absent. With questionnaires he found that when administrators and consultant agreed on the consultant's role, both rated highly the outcome of their consultations; when they disagreed they tended to rate their consultations less effective. This study supports those foregoing and includes a measure of effectiveness.

Lipham (17) did not ask his senior officials to make known their expectations for the principal's role upon which they presumably based their effectiveness ratings of 84 principals, but the fact that those rated "effective" conformed -- while those rated "ineffective" did not conform -- to the personality stereotype commonly held for the successful urban principal (e.g. high on: activity, drive, achievement dirve, upward mobility drive, social ability, feeling of security, emotional control), may be taken as evidence supporting the needs-expectations-effectiveness hypothesis. Or, perhaps, the "conformity" interpretation of it.

Although Chase, (9) it is true, did find a small positive relationship between teachers' stated level of job satisfaction and superintendents' ratings of their teacher-effectiveness, Brayfield and Crockett (7) report that between these two variables generally there is little relationship in

which much confidence can be placed. Their exhaustive report, which examines and summarizes the empirical literature bearing on the relationship between employee attitudes (satisfaction) and employee performance (Productivity), concludes:

> Let us make it clear that we expect the relation between satisfaction and job performance to be one of concomitant variation rather than cause and effect. (7, p. 416)

REFERENCES FOR CHAPTER III

(1) Andrews, J. H. M. "Administrative Significance of Psychological Differences between Secondary Teachers of Different Subject Matter Fields," Alberta Journal of Educational Research, 3:4 (Dec. 1957) 199-208

(2) Andrews, J. H. M. "A Deterrent to Harmony among Teachers," Administrator's Notebook, 4:7 (March 1958) 1-4

(3) Andrews, J. H. M. and Alan F. Brown, "Can Principals Exclude their Own Personality Characteristics When they Rate their Teachers?" Educational Administration and Supervision 45:4 (July 1959) 234-242

(4) Bidwell, C. E. "The Administrative Role and Satisfaction in Teaching," Journal of Educational Sociology, 29 (Sept. 1955) 41-47

(5) Bidwell, C. E. "Some Causes of Conflict and Tension among Teachers," Administrator's Notebook 4:7 (March 1956) 1-4

(6) Bidwell, C. E. "Administration and Teacher Satisfaction," Phi Delta Kappan, 37 (April 1956) 285-288

(7) Brayfield, A. H. and W. H. Crockett "Employee Attitudes and Employee Performance," Psychological Bulletin, 52:5 (1955) 397-424.

(8) Campbell, M. V. "Teacher-Principal Agreement on the Teacher Role," Administrator's Notebook, 7:6 (Feb. 1959) 1-4

(9) Chase, F. S. "Factors for Satisfaction in Teaching," Phi Delta Kappan, 33:3 (Nov. 1951) 127-132.

(10) Chase, F. S. "How to Meet Teachers' Expectations of Leadership," Administrator's Notebook, 1:9 (April 1953) 1-4.

(11) Congreve, W. J."Administrative Behavior and Staff Relations," Administrator's Notebook, 6:2 (Oct. 1957) 1-4.

(12) Ferneau, E. "Role-Expectations in Consultations," Unpublished Ph.D. Thesis, University of Chicago, 1954, reported in S. P. Hencley "Leader Behavior in Social Systems," Alberta Journal of Educational Research 7:2 (June 1961) p. 109.

(13) Getzels, J. W. and E. G. Guba "Role, Role Conflict and Effectiveness," American Sociological Review 19 (1954) 164-175.

(14) Getzels, J. W. and E. G. Guba "The Structure of Roles and Role Conflict in the Teaching Situation," Journal of Educational Sociology, 29 (Sept. 1955) 30-40.

(15) Griffiths, Daniel E. Administrative Theory. New York: Appleton-Century-Crofts, 1959.

(16) Guba, E. G. and C. E. Bidwell. Administrative Relationships. Chicago: Midwest Administration Centre, 1957.

(17) Lipham, J. M. "Personal Variables of Effective Administrators," Administrator's Notebook, 9:1 (Sept. 1960) 1-4

(18) Moyer, D. C. "Leadership that Teachers Want," Administrator's Notebook, 3:7 (March 1955) 1-4.

(19) Prince, Richard "Individual Values and Administrative Effectiveness," Administrator's Notebook, 6:4 (Dec. 1957) 14.

(20) Sharma, C. L. "Who Should Make What Decisions?" Administrator's Notebook, 3:8 (April 1955) 1-4.

(21) Sweitzer, R. E. "The Superintendent's Role in Improving Instruction," Administrator's Notebook, 6:8 (April 1958) 1-4.

(22) Valenti, J. J. "Measuring Educational Leadership Attitudes," Journal of Applied Psychology, 36 (Feb. 1952) 36-43

(23) Valenti, J. J. and T. E. Smith "Attitudes of Teachers and Administrators," Administrator's Notebook, 3:1 (Sept. 1954) 1-4.

CHAPTER IV

EXAMINATION AND EXTENSION

Stability and Fertility

Probably the most striking feature of Getzels' conceptualization of administrative relationships as a system of social action, viewed historically, has been its relative stability. Even though in the past decade, there has been tremendous change not only in the study of administration but in the social sciences in general, this model has withstood the test of time and remains basically in the original form advanced in 1952 (1, 3). A few of the studies reported did contain information that tended to indicate the need for a redevelopment of certain concepts of the theory but, with the exception of the addition of two "dimensions", the anthropological and the biological which were found neatly to fit the general structure of the model, no basic alteration has been found necessary. The possibility, therefore, is suggested that the period of agonizing adolescence through which a creative scientist guides his intellectual offspring is either unnecessary or yet to be faced. The view taken here, and supported in the following pages, is that the latter is so.

The second feature that strikes the student of administration theory development is the degree to which the original promise of generating research has been realized. A fertile

theory, such as this one would appear to be, promises to cause the development of research in greater amount, dispersion, and variety than has been witnessed with respect to Getzels' formulation. The possibility that this fact may be the reason for the relative stability noted in the opening paragraph cannot be overlooked; one way that theory develops is through a kind of "feedback" from the research it has generated. But the amount has not been large. The dispersion has been limited: even though the articulation among developers of administration research has been facilitated by the efforts of the NCPEA, the CPEA and UCEA, research support for Getzels' model has been concentrated chiefly at one centre -- the University of Chicago. The variety has similarly been limited; this is unfortunate because, to be judged sound, a theory should be capable of standing up under a variety of test situations. Those reported have been of the questionnaire-survey nature to a large degree; in decreasing numbers these were supplemented by or substituted with structured interviews, observations or card-sorts. Controlled experiments were conspicuous by their absence. That this may have been more by inclination than by accident could be taken as an inference of a statement included in one of Getzels' references to the 1954 study:

> . . .I should like to cite the investigation that Egon Guba and I did at Air University, Maxwell Air Force Base, several years ago since it deals simultaneously with several of the types of conflict derived from the model. *Moreover you will observe that it is not a laboratory test of the hypothetical formulations but a test in an ongoing educational situation.* (italics not in original) (2, p 162)

The concepts that are employed in the model are, nevertheless, very useful: they provide new insights into the dynamics of administrative relationships. In one sense, the present study was designed to strengthen some of the weaknesses of the model. It will analyze, redevelop and extend the theory to a point where it can more adequately conceptualize the interpersonal elements of supervision in both descriptive and predictive terms. It will extend the existing body of research in number, in dispersion, and in variety of procedures.

Limitations in the Personal Dimension

One basic weakness of the model is its psychological dimension. Specifically, the assumption that the need-disposition constitutes a necessary and sufficient analytic unit of personality commits the part-whole fallacy and ignores much of personality psychology since 1938.

For example, to Getzels as to Parsons a need-disposition is "the tendency to orient and act with respect to objects in certain manners and to expect certain consequences from these actions" (8, p. 114). Compare this with Murphy: ". . . a readiness to act in one way rather than in another . . . consisting, at least in part, of partial and finer symbolic acts" (6, p 980, adapted from H. B. English: A Student's Dictionary of Psychological Terms, 4th ed.), which is a standard definition of an attitude. Attitudes, for their part, are now customarily regarded as forming one of several analytic units -- and one of the less basic ones -- of personality. Such

overgeneralization is not, of course, unique. Reference has earlier been made to Henry A. Murray who, in 1938, built a personality theory around his concept of a need by expanding its definition to include virtually all drives, values, needs, attitudes and interests. The definition verbatim:

> A need is a construct (a convenient fiction or hypothetical concept) which stands as a force (the physico-chemical nature of which is unknown) in the brain region, a force which organizes perception, apperception, intellection, conation, and action in such a way as to transform in a certain direction an existing unsatisfying situation; a need is sometimes provoked directly by internal processes of a certain kind (viscerogenic, endocrinogenic, thalamicogenic) arising in the course of vital sequences, but more frequently (when in a state of readiness) by the occurrance of one of a few commonly effective press (or by anticipatory images of such press). Thus it manifests itself by leading the organism to search for or to avoid encountering or, when encountered, to attend and respond to certain kinds of press. (7, pp 123-124)

Murray goes on to define and specify the course of events resulting from this need which eventually lead to a situation that stills, appeases or satisfies the organism.

The point for the researcher is that the term need-disposition has become a hollow pot capable of accepting virtually any personality element or tendency he wishes to place inside it. Such versatility is not necessarily an asset. Lipham, for example, may better have been able to interpret his results had his psychological theory posited his personality variables (activity drive, achievement drive, upward mobility drive, social ability, feelings of security, and emotional control) as either (a) basic to personality or (b) those most

sensitive to the effects of role-personality conflict. Had he chosen the latter and attempted to predict the outcomes in terms of "sensitivity to conflict" or "anxiety" the interpretation of findings would have been much more refined in terms of differential effects among his principals. The same is generally true of several of the studies referred to in point form on pages 30 and 31 where predictions were made on the basis of generalized need-dispositions to generalized satisfaction, little or no account being taken of individual differences among teachers in their susceptibility or proneness to experience conflict, tensions, and dissatisfactions.

A more impressive example of this oversight appears in the report of the cornerstone Air University study where the data give every evidence of differences among officer-instructors in their emotionality, their tendency to experience different degrees of conflict from the same stimuli, and their sensitivity to civil-military conflict. The contribution of these data went virtually uninterpreted, particularly with respect to their effect on the structure and working of the model.

Perhaps a signal omission from the psychological "dimension" of the model stems from the age-old bifurcation of personality into the affective characteristics (e.g. attitudes, interests, needs, values) which _have_ been accepted as basic elements by personality theorists, and intellectual characteristics which in the past, have not. In short, intellectual ability or scholastic aptitute has either been outrageously subsumed

under the term "need-disposition" or has been omitted completely. Inasmuch as it has not been incorporated into the research designs, it probably has been omitted from the model in the same way that it is absent from several widely-accepted theories of personality, More recently, however, the practice has been to think in terms of the total personality, embracing the conative as well as the cognitive (and even, as with Eysenck's earlier publications, the physical) aspects.

Directional Nature of Predictions

The second basic weakness of the model, or the deductions that would appear to follow from it, lies in the unidirectional nature of its predictions. The greater the congruency the higher the effectiveness, conflict results in lower performance, satisfaction yields effectiveness. These statements are not inherent in the diagram, but are the deductions that customarily are made. The formula B=f(RxP) does not so state, neither does the structure of the theoretical formulation in diagram form. These state simply that behavior is a function of an interaction between internal and external forces. To learn how it functions implies a knowledge of all pertinent internal and external forces. A closer examination of internal forces may logically reveal that behavior may go either up or down as conflict increases, depending upon the nature of those forces in the individual under study. The notions being invoked here, of course, are those of Chapter One, or, more specifically, of stress as an energizer (4) and anxiety as a drive (5), and the differential effects of increasing

drive levels upon performance.

In summary, the Getzels model was called upon as a possible source of concepts to help answer the basic problem of this thesis: what happens to the effectiveness of a classroom teacher under conditions of interpersonal conflict created during the process of individual supervision, and why?

The model has helped fit the question into a rather secure framework encompassing both the individual teacher and the institution in which she is placed, a framework of organizational expectations and individual needs. The implications of such a conceptualization are that, where stress-inducing supervision can be thought of as resulting from the conflict of the teacher's personality elements and the demands of the institution as perceived by the teacher, less than optimum teaching effectiveness will be the result. But to the question "why?" the theory is less articulate. *Why*, for example, *should the perceived conflict be more shattering to one teacher's performance than to another's*? And *why*, once it is shattering, *does this affect* performance at all?

For answers to these two questions one must turn to a closer examination of the "personal" dimension. The first question calls for a discussion of individual differences in personal attributes and, perhaps more specifically, in self-concepts; the second requires an explanation of the dynamics of stress behavior.

REFERENCES FOR CHAPTER IV

(1) Getzels, J. W. "A Psycho-Sociological Framework for the Study of Administration," Harvard Educational Review, 22 (Fall 1952) 235-246.

(2) Getzels, J. W. "Administration as a Social Process," in A. W. Halpin (Ed.) Administrative Theory in Education. Chicago: Midwest Administration Centre, 1958.

(3) Getzels, J. W. and H. A. Thelen "The Classroom as a Unique Social System," The Dynamics of Instructional Groups: The 59th N.S.S.E. Yearbook, Part II. Chicago: The Society, 1960, pp. 53-82.

(4) McKinney, F. "Emotional Factors in Learning," Journal of General Psychology 9 (1933) 101-116.

(5) Mowrer, O. H. "A Stimulus-Response Analysis of Anxiety and its Role as a Reinforcing Agent," Psychological Review, 46 (1939) 553-565.

(6) Murphy, Gardner. Personality. New York: Harper and Bros. 1947.

(7) Murray, Henry A. Explorations in Personality. New York: Oxford University Press, 1938.

(8) Parsons, Talcott and Edward Shils. Toward a General Theory of Action. Cambridge: Harvard University Press, 1951.

CHAPTER V

THE PERSONAL DIMENSION

One need or many? One drive or many? One ability or several? The question of the one or the many, old as the ancient Greek philosophers, dogs inquiries into personality, motivation and intelligence. The parsimony of the one falls before the scientific demonstration of the many; the Balkanization of the many inhibits the pursuit toward unifying theory. And so it may be with the personal dimension of administrative theory in education.

Carl Rogers and Administrative Relationships

Although there are several theories in personality that posit a single need, drive or characteristic as the single basic unit for analyzing behavior, the choice of Carl Rogers' self-theory is neither fortuitous nor arbitrary. First of all, Rogers' basic need of self-actualization is sufficiently comprehensive to represent all those aspects that probably were implied in the original Getzels formulation. Without stretching definitions it is possible to substitute for the plural "need-dispositions" the Rogerian concept of personality with all its ramifications, the better to articulate Getzels' theory leads easily to a study of the dynamics of stress behavior arising out of role-personality conflict. Furthermore, Rogers has been able to claim extensive research support for his basic hypotheses (see, for example, 5); the major concepts of

self-theory have enjoyed widespread acceptance both in research and in clinical fields (see, for example, 2).

Carl Rogers' organismic theory of personality and behavior is most clearly stated in nineteen propositions. The propositions are the basic elements of his theory and have each been submitted to research for verification. To develop an adequate theory of teacher behavior under stress, relatively few of Rogers' propositions will be required. Because quoting out of context only those few that are required would impair their meaning, the complete list is presented verbatim:

> I. Every individual exists in a continually changing world of experience of which he is the centre. The private work may be called the phenomenal field, or the experiential field.
>
> II. The organism reacts to the field as it is experienced and perceived. This perceptual field is, for the individual "reality."
>
> III. The organism reacts as an organized whole to this phenomenal field.
>
> IV. The organism has one basic tendency and striving -- to actualize, maintain and enhance the experiencing organism. Rather than many needs and motives, it seems entirely possible that all organic and psychological needs may be described as partial aspects of this one fundamental need.
>
> V. Behavior is basically the goal-directed attempt of the organism to satisfy its needs as experienced in the field as perceived.
>
> VI. Emotion accompanies and, in general, facilitates such goal-directed behavior, . . . the intensity of the emotion being related to the perceived significance of the behavior for the maintenance and enhancement of the organism.
>
> VII. The best vantage point for understanding behavior is from the internal frame of reference for the individual himself.

VIII. A portion of the total perceptual field gradually becomes differentiated as the self. Though some authors use the term "self" as synonymous with "organism" it is here being used in a more restricted sense, namely, the awareness of being, of functioning.

IX. As a result of interaction with the environment, and particularly as a result of evaluational interaction with others, the structure of self is formed -- an organized, fluid but consistent conceptual pattern of perceptions of characteristics of the "I" or of the "me", together with values attached to these concepts.

X. The values attached to experiences and the values which are a part of the self-structure, in some instances, are values experienced directly by the organism, and in some instances are values introjected or taken over from others, but perceived in distorted fashion, as if they had been experienced directly.

XI. As experiences occur in the life of the individual, they are either (a) symbolized, perceived and organized into some relationship to the self, (b) ignored because there is no perceived relationship to the self-structure, (c) denied symbolization or given distorted symbolization because the experience is incomsistent with the structure of self.

XII. Most of the ways of behaving which are adopted by the organism are those which are consistent with the concept of self.

XIII. Behavior may, in some instances, be brought about by organic experiences and needs which have not been symbolized. Such behavior may be inconsistent with the structure of self, but in such instances the behavior is not "owned" by the individual.

XIV. Psychological maladjustment exists when the organism denies to awareness significant sensory and visceral experiences, which consequently are not symbolized and organized into the gestalt of the self-structure. When this situation exists there is a basic or potential psychological tension.

XV. Psychological adjustment exists when the concept of the self is such that all the sensory and visceral experiences are, or may be, assimilated on a symbolic level into a consistent relationship with the concept of self.

XVI. Any experience which is inconsistent with the organization or structure of self may be perceived as a threat, and the more of these perceptions there are the more rigidly the self-structure is organized to maintain itself.

XVII. Under certain conditions, involving primarily complete absence of any threat to the self-structure, experiences which are inconsistent with it may be perceived, and examined, and the structure of self revised to assimilate and include such experiences.

XVIII. When the individual perceives and accepts into one consistent and integrated system all his sensory and visceral experiences, then he is necessarily more understanding of others, and is more accepting of others as other individuals.

XIX. As the individual perceives and accepts into his self-structure more of his organic experiences, he finds that he is replacing his present value-system -- based so largely upon introjections which have been distortedly symbolized -- with a continuing organismic valuing process. (4, pp 481-533)

The pertinence of self-theory to the problem of this thesis, as well as to the problems raised by the Getzels model, is seen by attending specifically to Propositions IV, V, VI, IX, XIV, and XVI. Rogers claims that a person seeks those activities (or attitudes or behaviors generally) that tend to maintain and enhance the concept of self, and that this is the fundamental need for self-actualization (IV,V). This notion has support from the Canadian psychologist, Hebb, who tends to prefer a single general drive theory of motivation accompanied by specific drives for corresponding to primitive biological needs (3, pp. 157-158). Rogers also holds that this self-concept is built through evaluational interations (IX), quite similar to the interpersonal theory of "reflected appraisals from significant others" of Harry Stack Sullivan (6) which has received

experimental support from the present writer (1). These evaluational interactions would include, among other things, such events as situations of individual supervision as well as experiences of success and failure in earlier cognitive tasks, e.g. schoolwork.

Rogers' propositions include mention (VI) that emotion attaches to behavior in proportion to its perceived significance to the self-concept that has been built up. One could readily refer this to the supervision wituation where the purpose of the event was the supervisor's evaluation of the teacher's behavior; in such a situation there is every liklihood that a teacher will react with some emotion to the supervisor's stated or implied comparison of her own behavior with the standards or expectations he holds for her teaching. Furthermore, when the situation is inconsistent with the self-concept, adjustment is difficult and psychological tension results (XIV). This is strikingly similar to the statement that when the individual perceives organizational expectations to be incongruent with his personal needs (now expanded to include, or be substituted by, Rogers' "need"), there is role-personality conflict and satisfaction is low. Indeed, the correspondance between those propositions cited and the statement od administrative relations as a social system is sufficient that the need for this psychological sojourn may be questioned. Briefly, it served the purpose of providing a broader conceptualization of the "personal dimension", its dynamic development and probable

consequences under certain conditions, particularly those that may be construed as ego-threatening or inconsistent with the total concept of self. It did not, of course, specify the precise results of potential conflict on actual performance.

It is interesting to note that Proposition XVII implies that individual supervision in which there is no perceived threat, or in which needs-expectations congruence is high, results in a constructive re-structuring of the self-concept. In a sense, this is a major goal of all supervisors, the way to which is clear: without the use of threat, real or imagined, to draw organizational demands and individual personalities more in line with one another.

There are several reasons why a teacher may perceive a supervisor's comment as ego-threatening depending, it would appear, upon the nature of the self-image the teacher has built up over years of evaluational interactions, both immediate and indirect. One logical deduction would lead to the hypothesis that a person who has built a self-concept of success in cognitive or scholastic tasks would so perceive an event that challenged this self-picture, i.e. which revealed failure in such a task or related activity. For example, a person of high general ability throughout schoolwork and academic endeavours holds certain self-expectations when confronted with a new task that differ from one of low general ability. Whether or not they actually think they are "clever" or that they have a high "I.Q." is quite beside the point; they come

to expect to attain whatever success is consistent with the past, and this to be reflected in evaluational interactions (actual "marks" or remarks) with others. One would, therefore expect this person to experience difficulty in assimilating into his self-image an evaluation indicating personal failure; it would be ego-threatening and, perhaps depending upon the lability of the person, stressful. With all else being equal (particularly with regard to lability), the person who has experienced repeated failure would have built up a self-concept accordingly, would more readily assimilate a failure situation into his self-structure, and would be less likely to experience psychological stress from it. Such a person we have conditioned to failure.

The Variable of "Scholastic Aptitude"

The writings of Carl Rogers do not display a particular concern with the cognitive aspects of personality. They do, however, indicate a concern for the total personality of which cognitive ability is a part. The logic of the preceding section concluded that there is an important relationship between measured scholastic aptitude or general intelligence on the one hand, and self-expectations of success or failure on cognitive tasks on the other. The relationship probably works both ways: a child learns that he has done well on a test and incorporates this into his self-image; a child expects to do poorly on a task and this contributes to his probably failure. There has been evidence of spelling improvement following

counselling which had the effect of altering the student's earlier self-concept of being a poor speller.

It is here suggested that measured scholastic aptitude (or, as it may be commonly termed, general intelligence) is an indicator of performance under conditions of failure-stress. This suggestion is predicated on the notion that measured scholastic aptitude is a composite of "innate ability" and past experience, and that this past experience has contributed heavily to the development of the self-image regarding success of performance in complex tasks. The variable with which we are concerned is, therefore, scholastic aptitude. The justification for using this variable, however, depends upon a dynamic interpretation of an intelligence test score.

REFERENCES FOR CHAPTER V

(1) Brown Alan F. "The Self in Interpersonal Theory: The Relation between Perceptions of Self and of Significant Others," Alberta Journal of Educational Research, 3:3 (Sept. 1957) 138-148.

(2) Hall, C. S. and G. Lindzey. Theories of Personality. New York: Wiley, 1957.

(3) Hebb, D. O. A Textbook of Psychology. Philadelphia: W. B. Saunders, 1959.

(4) Rogers, Carl R. Client-Centred Therapy. Boston: Houghton-Mifflin, 1951.

(5) Rogers, Carl R. and R. F. Dymond. Psychotherapy and Personality Change. Chicago: University of Chicago Press, 1954.

(6) Sullivan, Harry Stack. Conceptions of Modern Psychiatry. Washington: William Alanson White Foundation, 1947.

CHAPTER VI

DEFINING ANXIETY, NEUROTICISM, STRESS AND DRIVE

Usage in Technical Literature

Within the general problem area of the effects of psychological stress upon task performance, definitions have frequently been conflicting. There is evidence of a number of terms referring to the same general phenomena, sometimes interchangeably: anxiety, stress, frustration, tension, emotional tension, neuroticism, arousibility, fear and threat. Writers frequently use one of these terms as the central concept of their work; the fact that their central concepts show more consistency from one to another than do the actual terms employed is a source of confusion within the literature. Anxiety is the central concept to Taylor (49), neuroticism is to Eysenck (17) even though the title of his most recent book on neuroticism and extraversion uses the terms "Anxiety and Hysteria". Janis (22) regards emotional tension as the generic term from which the others are derived; Murphy (35) calls it tension; and Symonds (47) seems to prefer reference to stress.

Symonds differentiates among some of these terms in the following manner: frustration is interference with the satisfaction of some need; anxiety is the emotional reaction aroused by anticipated frustration or punishment; tension arises from continuous frustration or threat thereof and

accompanies anxiety and has a physiological base; stress is a condition or situation or stimulus that arouses anxiety and produces tension.

To Murphy, anxiety is an emotional state arising when a continuously strong desire seems likely to miss its goal (that is, it is need-frustration, actual or anticipated). Tension includes all the energies concentrated in the various parts of the body.

May (33) finds that the concensus of several investigators in the field of anxiety (particularly Mowrer, Horney and Sullivan) is that frustration *per se* does not cause conflict:

> The basic question is, rather, what essential value is threatened by the frustration?. . . The problem is not the frustration itself, but whether the frustration threatens some mode of interpersonal relationships which the individual holds vital to his security and self-esteem. (33, p. 211)

Janis (22) tends to suggest that the empirical approach will result in a common ground for research workers, a ground which he feels is yet lacking. The events referred to as psychological stress would, therefore, be those changes in the environment which *typically* induce a high degree of emotional tension (including anxiety) and which interfere with normal response patterns. Hence one would speak of a stress situation whether or not it actually produced stress reactions in a particular individual.

The man whose name is virtually synonymous with the concept of stress, Hans Selye (42, 43), provides the following definition:

> Stress is the state manifested by a specific syndrome which consists of all the nonspecifically induced changes within a biologic system. (43, p. 54)
>
> . . . the whole development of the (stress) reaction largely depends on conditioning factors. These can be invariables which act upon us from within . . . (internal conditioning), as well as variables which influence our body simultaneously with the agent from without (external conditioning). All these are integral elements of the response during stress. . . . (43, p. 98)

Selye makes it clear throughout that emotional tension *is* one such nonspecific, and very powerful, stress agent. Another significant concept is that there are great individual differences among persons (and other animals) in their susceptibility to stressful states. Such concepts as these are the result of long series of experiments, mostly at Montreal. In a philosophic mood, however, Selye draws the implication that "stress plays a role in such diverse manifestations of life as aging, the development of individuality, the need for self-expression (*sic*), and the formulation of man's ultimate aims." (43, p. 253). These, of course, he claims to have logically deduced from his medical research, one generalization of which is that stress is usually the outcome of a struggle for the self-preservation of parts within a whole, whether of cells within man, of man within society, or of species within the whole animate world.

Because neuroticism is a central theme of Eysenck's writing, (the other one is extraversion-introversion), one would expect a definition to be found. In fact, operational definitions of neuroticism (i.e. in terms of tests or other indications of behavior) are much more in evidence than any precise Aristotelian definition. Neuroticism is, however, clearly the opposite end of a continuum from emotional stability (13); it is the equivalent of 'badly organized personality' (12); or 'lack of balanced personality arising from emotional causes' (14). It can be used interchangeably with 'anxiety' (16), and is much the same thing as 'arousibility' or emotionality (17) indicating 'emotional overresponsiveness, and liability to neurotic breakdown under stress' (18). Although the term "neuroticism," like several others in this section, may vary in its meaning from that of everyday parlance, the attempt in this thesis has been to remain as consistent as possible with technical usage. Thus, Eysenck (15) justifies the use of the term neuroticism to refer to any point on a continuum denoting a degree less than normal, to normal, to a degree that is greater than normal, on the basis of his hierarchical theory of personality. The difference between a normal and a neurotic is one of degree, not of kind (a psychotic shows a difference of kind), so that there need be no hesitation in referring to normals as possessing various degrees of neuroticism.

May (33) draws a useful distinction between normal anxiety and neurotic anxiety. Normal anxieties are based on

rational fears and customarily are situational even though some are long-lived. They do not, as such, have anything to do with neuroticism, or to the emotionality and irrationality of the subject. Neurotic anxieties, on the other hand, are based on irrational fears or fears that once were rationally based but, presumably because of neurotic tendencies in the individual, are now out of proportion to reality. Neurotic anxieties are, therefore, deep-seated and persistent although they need not be so irrational or fearsome as to require psychological or medical attention. Hence, neurotic anxieties follow a continuum that includes normals. An additional ambiguity in the term anxiety is that it is often used commonly to denote 'anxiety state', or a behavior disorder requiring attention.

This thesis deals with normals. Thus it is differences in 'neuroticism' or 'neurotic anxiety' as a personality characteristic that is the subject for study. The discussion that follows will draw upon writings in the field of neuroticism, and in anxiety when it appears to this investigator that neurotic anxiety is the concept being employed.

The terms fear and threat, while not in such common technical usage, often appear in the literature. To Murphy (35), fear is an emotion marked by alarm or dread; it is sometimes used where the more appropriate term would be anxiety. Janis (22) notes that threat occurs when cues heralding oncoming danger occur without any noxious stimuli being actually present although the difference between apprehension concening anticipated danger on one hand, and emotional reactions to actual

stress on the other, may be purely quantitative.

All of the terms used thus far share at least one common feature: each has been shown at one time or another to have the properties of a drive. Hebb (21) defines a drive as:

> some process which provides the energy of movement but which, like the engine of an automobile, does not determine what the movement will be. Stimuli then have two different roles in behavior. One, which we shall call the arousal function, is to make the animal active, or to tend to do so, apart from determining what form his activity will take. The other, the cue function, guides the response. . . . Thus activity in the arousal system has the essential characteristics of a drive. . . . (21, pp. 157-158)

Although the question of whether there is a single drive or several remains open, Hebb favours the notion that there is a single generalized drive mechanism which can be excited by different conditions, yet retains the notion of separate drives for biologically primitive needs each with its own pattern of activity in the brain-stem arousal system but overlapping.

Classification of Stress Stimuli

A classification of stress situations has been proposed by Schwab and Pritchard (41), being the one which they used in their research: (a) mild stress, the effects of which last from seconds to hours and which include annoying insects, public appearances and missing a train; (b) moderate stress, the effects of which last hours to days, including periods of overwork, gastric upset, unwelcome guest, absence of a loved one; and (c) severe stress, lasting from weeks to years, including prolonged separation from family, death of a loved one, financial losses, illnesses, operations and disasters. This classification

is, of course, nothing more than a statement that stress intensity and duration are mutually related. The differences are of degree.

Another classification of stress situations that has been used in a number of research studies is based on types of stress. Lazarus *et al*. (23) refer to failure-induced stress and task-induced (i.e. distraction) stress. Stopol (45) found that the effects of these two types of stress vary independently although, since he did not provide a metric for stress intensity, one cannot be sure that independent variance was due to type rather than to intensity. Broverman and Lazarus (3) revised the concept and referred to their two types of stress as ego-threatening and cognitive-interference stress respectively. While the earlier Lazarus terms are more operationally suited to research the later ones, denoting the same conditions, are more precise in describing the stress effects upon the individual. The Lazarus classification would cut across the classification provided by Schwab and Pritchard; there are presumably, degrees of intensity and duration in each type of stress. The essence of the Lazarus classification seems to be that whereas one type is ego-centred the other is task-centred. Ego-threatening stress may inhibit the individual's ability to perform during the conduct of the task; cognitive-interference stress may inhibit the possibilities of anyone performing the task. In the final analysis, of course, it is possible to construe any stress situation as ego-threatening depending upon the stress-proneness of the individual personality. The

conflict arising out of the needs-expectations incongruence in the Getzels model could be so defined.

The Factors and Correlates of Anxiety and Stress

O'Connor *et al.* (36), in a factor analysis of manifest anxiety (the Taylor MAS), found these five component factors: chronic anxiety or worry, increased physiological reactivity, (e.g. blush or sweat) to emotional stimuli, sleep disturbances associated with inner strain, feeling of personal inadequacy or lack of self-confidence, and a motor factor of tension, restlessness or nervousness. Wilensky (50) factor analyzed ten variables associated with anxiety and report two factors: subjectively-reported experienced anxiety, and physical conditions and reactions. Buss *et al.* (4) defined anxiety operationally as a result of an investigation that revealed nine components: clinically observed distractibility, restlessness and physiological concomitants, subjectively reported feelings of tenseness, worry-apprehension, somatic complaints, physiological concomitants, muscular tension and overall feeling of anxiety. Elizur (11) found high inter-rater agreement on manifest anxiety when the term was defined as an inner state of insecurity which may take one or more of the following forms: fears, phobias, lack of self-confidence, extreme shyness, ideas of reference, and marked sensitivity.

The place of anxiety in motivation has been studied by Lindsey (25) who employed anxiety as an intervening variable operating between motivation and behavior. Its relation to the

achievement motive (n Achievement) has been carefully studied by McClelland (27), McClelland and Apicella (28), Raphelson (37), Raphelson and Moulton (38), and Sarason and Mandler (31, 39, 40). The general findings are that anxiety is negatively correlated with n Achievement, that high failure-stress (ego-threatening) results in poorer cognitive performance and produces greater variance in results than does low failure-stress. Further, n Achievement appears directly related to physiological tension. As expectation of failure increases (ego-threat stress stimuli increase) there is a decrement in test performance (complex task) of high-anxiety (old or persistent level) subjects. Uncertainty as to success or failure may cause withdrawl in high-anxiety subjects submitted to intense ego-involvement. As early as 1933, McKinney (30) found that mild anxiety acted as an exergizer but that severe levels thwarted or redirected motivation.

The effect of stress upon cognitive tasks of the intelligence test sort has been studied by Lazarus and Eriksen (24), Moldawsky and Moldawsky (34), Bevan and Meier (2), Tallarico and Reitman (48), Matarazzo and Phillips (32), and Deese, Lazarus and Keenan (9). An induced failure-stress group appears to produce more variance (the high's get higher, the low's lower) than do a no-stress group; the former also produce more errors which are partly compensated by an increase in speed over the latter. Digit-span scores are more sensitive to situational anxiety, particularly experimentally-induced ego-threat, than are scores from higher-level mental processes

of vocabulary and digit symbol. Its effects upon anagrams-type tasks are equivocal. Generally speaking, anxiety is established as a very powerful factor in learning processes and in the performance of complex tasks when other factors, especially those in the cognitive field, are held constant.

The correlates of anxiety level with other personality variables have shown problematical results. When pencil-and-paper tests are used, criticism runs in both directions: a low correlation is blamed on the crudity of the instruments, a high correlation is attributed to the common paper-and-pencil factor. Silverman (44) found generally low (less than .36) correlations between anxiety scales and each of the fifteen manifest needs of the Edwards Personal Preference Schedule. Bendig found a correlation of .77 between anxiety (Taylor MAS) and neuroticism (Eysenck MPI-N) and of -.35 between anxiety and extraversion-introversion (MPI-E). Cowen *et al.* (8) found that low-anxiety subjects have more adequate self-regarding concepts (Bills ' Index of Adjustment and Values) than those with high anxiety but that anxiety level bore no relation to behavioral rigidity in perceptual tasks (Luchins water-jar). Earlier, however, Cowen (6) found the mild-stress group less rigid than the severe stress group in a problem-solving situation; the severe stress group also showed slower response to testing. Lindzey and Newburg (25) found anxious subjects revealed themselves to be rigid and restricted, vague and hesitant in their TAT protocols. Gorlow *et al.* (20) showed significant differences between anxious and non-anxious subjects on the

Rorschach, but Stopol (45) found that experimentally-induced stress failed to have any effect upon Rorschach protocols. That is, old and persistent anxiety levels that show on anxiety tests tend to be reflected in projective test performance whereas situational stress does not.

While most of the reported studies were carried out for the specific purpose of investigating the effects of anxiety and stress, usually within a stated theory of behavior, in some others there has been no concealing of efforts to produce a convenient and efficient measure of anxiety. Freeman (19), in his "Suggestions for a standardized 'stress' test," advised:

> While we may ultimately find a physiological parameter . . . that will predict the degree of emotional stability possessed by the individual, that day is far distant and the approach to it would seem to come through the study of *voluntary action under experimentally induced* stress. (italics not in original) (19)

This counsel was, in general, heeded. The following decade witnessed subjects being submitted to a dazzling array of treatments: electronically timed flashing lights, bells, buzzers, electric shock, air-blasts to the cornea, verbal heckling, falsified (low) scores, withdrawn rewards, threat of punishment or failure, all manner of ego-threat and downright berbal abuse -- all in the quest for stress. At least two studies (46, 5) have used McClelland and Apicella's (28) classification of comments ingeneously devised to induce failure-stress. In this method the experimenter:

(a) calls attention to S failure (actual or fabricated)

(b) compared S to his disadvantage to the group

(c) complains S is ruining the experiment, and

(d) becomes aggressive toward S, attaching him verbally

Indeed, experimenters have been so successful in inducing stress reactions that Cowen (6) carefully described the steps he took to extinguish the condition at the close of the experiment.

Accepted Definitions of Terms Used in this Thesis

This chapter developed the meanings of several central concepts used in this thesis. The following list of accepted definitions is taken directly from the standard reference, English and Engligh: A Comprehensive Dictionary of Psychological and Psycho-Analytic Terms. In each instance, the writer has indicated the agreement between the lexicographers' definition and one that is accepted in this thesis, following the word "Note".

Achievement Drive: the tendency to work with energy and persistence at something deemed important; ambition manifested in action (Note: this agrees with Edwards, and is accepted here.)

Anxiety: 1. an unpleasant emotional state in which a present and continuing strong desire or drive seems likely to miss its goal; 2. a secondary drive, the symptom of which is that the stimulation of the anxiety response depresses the rate of the responses usual in the situation and produces other behaviors inappropriate to the situation. (Note: 1 is similar to Murphy and Symonds, 2 to Spence and Taylor; they are not antithetical; both are accepted here.)

Drive: (or Drive State, or D): a tendency, initiated by shifts in physiological balance, to be sensitive to stimuli of a certain class and to respond in any of a variety of ways that are related to the attainment of a certain goal....

The drive state may be manipulated by creating certain conditions which may be called drive arousal or drive stimulus. (Note: this agrees with Hebb, Spence and Farber, and subsumes Selye's concept of stress; it is accepted here.)

Need: the lack of something which, if present, would tend to further the welfare of the organism . . . or to facilitate its usual behavior; or the thing, activity or condition that is lacking. (Note: this does not agree with 'need-disposition' in Getzels, nor with Murray's 'need'; it agrees with Edwards' use of the term 'need for achievement' although that may better be called 'achievement drive' as above. Little use will be made of the term by itself.)

Neurotic: characterizing behavior that resembles that of neurosies; erratic and excessively emotional behavior (Note: see next entry.)

Neurotic Character: the basic quality, or complex of qualities or traits, that leads to the overt manifestation of neurosis; the neurotic character may be conceived as a matter of degree. (Note: there is no essential disagreement between these two definitions and Eysenck's concept of neuroticism which is accepted here. Eysenck believes, for example, that the difference between normals and neurotics is one of degree of the 'neuroticism' variable. The relation between the lexicographers' definitions of anxiety and neurotic is noteworthy.)

Personality: (the lexicographers list, and criticize, several meanings but seem to favour Allport's definition as the best common denominator; it follows.) the dynamic organization within the individual of these psychophysical systems that determine his unique adjustment to his environment. (Note: the writer agrees with English and English.)

Stress: a force, applied to a system, sufficient to cause strain or distortion in the system. (Note: Selye would agree, but would probably disagree on what is distortion and what is not. Selye would expand it to include all nonspecifically induced changes. He would also question the notion that stress is a force; rather, it is the state, the stress agent is the force. Whether one speaks psychologically or physiologically it is difficult to equate stressor with stress. The writer accepts Selye's definition, stated earlier in the chapter, but reconciles it partly with English and English by noting that infinitessimal amounts of stress are probably not worthy of our attention here. But English and English's definition, confuse stress and drive.)

Stress Test: a test in which the task itself, together with the prescribed conditions of testing, puts the subject under great emotional strain; the tasks are generally lifelike in nature and the purpose of the test is partially or wholly disguised. (Note: this agrees with the writer's usage throughout, although to avoid confusion with pencil-and-paper tests it is normally referred to as the stress stimulus or stress conditions meaning that the stress test creates the state of stress in the subject. The writer does not see the necessity of the adjective 'great' in the definition, however; the amount of strain is herein believed to vary.)

REFERENCES FOR CHAPTER VI

(1) Bendig, A. W. "Extraversion, Neuroticism and Manifest Anxiety," Journal of Consulting Psychology 21:5 (Oct. 1957) 398.

(2) Bevan, W. and R. A. Maier "Emotional Tension and the Generality of its Effect upon Intellectual Performance," Journal of Personality, 26:3 (Sept. 1958) 330-336.

(3) Broverman, D. M. and R. S. Lazarus "Individual Differences in Task Performance under Conditions of Cognitive Interference," Journal of Personality 26:1 (March 1958) 94-105.

(4) Buss, A. H., M. Wiener, A. Durkee and M. Baer "The Measurement of Anxiety in Clinical Situations," Journal of Consulting Psychology, 19:2 (April 1955) 125-129.

(5) Charles, H. and E. Mech "Performance in an Operational 'Stress' Situation Related to a Projective Technique," Journal of Educational Research, 48 (March 1955) 509-519.

(6) Cowen, E. L. "The Influence of Varying Degrees of Psychological Stress on Problem-Solving Rigidity," Journal of Abnormal and Social Psychology, 47 (1952) 512-519.

(7) Cowen, E. L., M. Weiner and J. Hess "Generalization of Problem-Solving Rigidity," Journal of Consulting Psychology, 17:2 (1953) 100-103.

(8) Cowen, E. L., F. Heilizer, H. L. Axelrod and S. Alexander "The Correlates of Manifest Anxiety in Perceptual Reactivity, Rigidity and Self-Concept," Journal of Consulting Psychology, 21:5 (Oct. 1957) 405-411.

(9) Deese, J., R. S. Lazarus, and J. Keenan "Anxiety, Anxiety Reduction, and Stress in Learning," Journal of Experimental Psychology, 46 (July 1953) 55-60.

(10) Edwards, Allan L. Personal Preference Schedule and Manual, (Revised Ed.). New York: Psychological Corp; 1959.

(11) Elizur, A. "A Content Analysis of the Rorschach with Regard to Anxiety and Hostility," Journal of Projective Techniques, 13 (1949) 247-283.

(12) Eysenck, H. J. Dimensions of Personality. London: Routledge and Kegan Paul, 1947.

(13) Eysenck, H. J. The Scientific Study of Personality. London: Routledge and Kegan Paul, 1952.

(14) Eysenck, H. J. The Structure of Human Personality. London: Methuen, 1953.

(15) Eysenck, H. J. "A Dynamic Theory of Anxiety and Hysteria," Journal of Mental Science, 101 (Jan. 1955) 28-51.

(16) Eysenck, H. J. "Reminiscence, Drive and Personality Theory," Journal of Abnormal and Social Psychology, 53 (1956) 328-333.

(17) Eysenck, H. J. The Dynamics of Anxiety and Hysteria. London: Routledge and Kegan Paul, 1957.

(18) Eysenck, H. J. The Maudsley Personality Inventory and Manual. London: University of London Press, 1959.

(19) Freeman, G. L. "Suggestions for a Standardized 'Stress' Test," Journal of General Psychology, 32:1 (Jan. 1945) 3-11.

(20) Gorlow, L., C. N. Zimet and H. J. Fine "The Validity of Anxiety and Hostility Rorschach Content Scores Among Adolescents," Journal of Consulting Psychology, 16 (Feb. 1952) 13-75.

(21) Hebb, Donald O. A Textbook of Psychology. Philadelphia: W. B. Saunders and Co., 1959.

(22) Janis, I. L. Psychological Stress. New York: Wiley and Sons, 1958.

(23) Lazarus, R. S., J. Deese, and R. Hamilton "Anxiety and Stress in Learning: The Role of Intraserial Duplication," Journal of Experimental Psychology, 47 (1954) 111-114.

(24) Lazarus R. D. and C. W. Eriksen "Effects of Failure Stress Upon Skilled Performance," Journal of Experimental Psychology, 43 (Feb. 1952) 100-105.

(25) Lindzey, G. and A. S. Newburg "Thematic Apperception Test: A Tentative Appraisal of Some 'Signs' of Anxiety," Journal of Consulting Psychology, 18 (Dec. 1954) 389-395.

(26) McClelland, D. C. Personality. New York: Dryden Press, 1953.

(27) McClelland, D. C. Studies in Motivation. New York: Appleton-Century Crofts, 1955.

(28) McClelland, D. C. and F. S. Apicella "A Functional Classification of Verbal Reactions to Experimentally Induced Failure," Journal of Abnormal and Social Psychology, 40 (1945) 376-490.

(29) McClelland, D. C., J. W. Atkinson, R. A. Clark and E. L. Lowell. The Achievement Motive. New York: Appleton-Century-Crofts, 1953.

(30) McKinney, F. "Emotional Factors in Learning," Journal of General Psychology, 9 (1933) 101-116.

(31) Mandler, G. and S. B. Sarason "A Study of Anxiety and Learning," Journal of Abnormal and Social Psychology, 47 (1952) 166-173.

(32) Matarazzo, J. D. and J. S. Phillips "Digit Symbol Performance as a Function of Increasing Levels of Anxiety," Journal of Consulting Psychology, 19:2 (April 1955) 131-134.

(33) May, Rollo. The Meaning of Anxiety. New York: Ronald, 1950.

(34) Moldawsky, S. and P. C. Moldawsky "Digit Span as an Anxiety Indication," Journal of Consulting Psychology, 16 (April 1952) 115-118.

(35) Murphy, Gardner. Personality. New York: Harper and Bros., 1947.

(36) O'Connor, J. P., M. Lorr and J. W. Stafford "Some Patterns of Manifest Anxiety," Journal of Clinical Psychology, 12:2 (April 1956) 160-163.

(37) Raphelson, A. C. "The Relationships Among Imaginative, Direct Verbal, and Physiological Measures of Anxiety in an Achievement Situation," Journal of Abnormal and Social Psychology, 54 (1957) 13-18.

(38) Raphelson, A. C. and R. W. Moulton "The Relationship between Imaginative and Direct Verbal Measures of Test Anxiety under Two Conditions of Uncertainty," Journal of Personality, 26:4 (Dec. 1958) 556-567.

(39) Sarason, S. B. "Effect of Anxiety, Motivational Instruction and Failure on Serial Learning," Journal of Experimental Psychology, 51 (1956) 253-260.

(40) Sarason, S. B., G. Mandler, and P. C. Craighill "The Effect of Differential Instructions on Anxiety and Learning," Journal of Abnormal and Social Psychology, 47 (1952) 561-566.

(41) Schwab, R. and J. Pritchard "Situational Stresses and Extrapyramidal Disease in Different Personalities," Life Stress and Bodily Disease, Association for Research in Nervous Mental Disease, 1949, pp. 48-60.

(42) Selye, Hans. The Story of the Adaptation Syndrome. Montreal: Acta, Inc., 1952.

(43) Selye, Hans. The Stress of Life. Toronto: McFraw-Hill, 1956.

(44) Silverman, R. E. "The Edwards Personal Preference Schedule and Social Desirability," Journal of Consulting Psychology, 21:5 (Oct. 1957) 402-404.

(45) Stopol, M. S. "Rorschach Performance in Relation to Two Types of Stress," Journal of Consulting Psychology, 18 (1954) 11-15.

(46) Stopol, M. S. "The Consistency of Stress Tolerance," Journal of Personality, 23 (1954) 13-29.

(47) Symonds, P. M. "What Education Has to Learn from Psychology: VI. Emotion and Learning," Teachers College Record, 60:1 (Oct. 1958) 9-22.

(48) Tallarico, R. B. and E. E. Reitman "Anagram-Solving Ability as an Index of Anxiety," Journal of Clinical Psychology, 15:2 (April 1959) 181.

(49) Taylor, J. A. "Drive Theory and Manifest Anxiety," Psychological Bulletin, 53 (1956) 303-320.

(50) Wilensky, H. "A Factorial Study of Behavioral and Psychological Measures of Anxiety," Journal of Consulting Psychology, 21:3 (June 1957) 216.

CHAPTER VII

THEORETICAL IMPLICATIONS OF STRESS-REACTIONS STUDIES

Drive Levels and Task Performance

Anxiety has been regarded as a drive with respect to behavior in the realm of learning. In fact, it is normally regarded as a very potent drive. In the previous chapter, reference was made to studies that found anxiety, when cognitive factors were held constant, to be a strong factor in learning or in level of performance in complex tasks. At the close of Chapter Five reference was made to Hebb who defines drive as some process which provides non-directed energy, and to his conclusion that activity in the arousal system of man was found to have the essential characteristics of a drive.

Eysenck's inclusion of the word "arousibility" in his definition of neuroticism reveals his use of the term to connote drive properties. Selye's long experiments with the general adaptation syndrome (24) led to the conclusion that stress is a state induced in the organism by some nonspecific stimulus, including emotional tension, which serves the purpose of readying the organism for action. Hence, emotional tension and its resultant state of stress must be thought of drive-producing.

In Hull's theory of learning, a variable is considered to have the properties of a drive if:

> (a) the elimination or reduction in the magnitude of the variable is reinforcing, i.e. leads to the increased probability of recurrence in the same situation of the responses that precede the modification of the variable, and/or (b) the presence of the variable energizes or intensifies whatever reaction tendencies exist in the given situation. (12)

In a famous paper entitled "A stimulus-response analysis of anxiety and its role as a reinforcing agent," Mowrer (21) has put the notion of anxiety as a drive on a firm basis; experimental work by him and by others has served to confirm this notion.

It would seem to follow then that one's susceptibility to stress or emotional tension (e.g. as indicated by degree of anxiety or neuroticism must be regarded as part of the general drive level of the individual, and it would also follow that individual differences in anxiety or neuroticism will enter into the general Hullian formula for increasing the amount of drive present in any task situation. This is at least in part the argument that Spence, Taylor and others of the Iowa group advance (26, 27, 28, 29, 30). For example, Taylor states that, according to Hull, all habits activited in a given situation combine multiplicatively with the total effective drive state operating at the moment to form excitatory potential. This is expressed in the formula E = f(HxD). Total effective drive (the D in the formula) was regarded as determined by the summation of all extant need states. Taylor and others have attempted to demonstrate this with respect to simple conditioning, contrasting low and high anxiety groups. All reported studies

have shown the high-anxiety subjects to have a higher rate of conditioning (conditioning the eye-blink by using an air-blast to the cornea as the UCS and a pure tone as the CS) than the low-anxiety subjects (25, 26, 27, 28, 29).

On the other hand, Eysenck (10, 11) has shown (1) that their anxiety scale (Taylor MAS) has projections on two dimensions, neuroticism (which is highly correlated with anxiety) and extraversion-introversion (which is not, and (2) that conditionability (as distinct from learning on complex tasks) is more related to the latter dimension than to the former. In other words, the question can be raised as to whether the Iowa eye-blink was more determined by extraversion-introversion than by anxiety. Eysenck concluded that: ". . . there is no reason to assume that anxiety-neuroticism _as an irrelevant drive_ should produce any increase in the rate of conditioning" (11, p. 93).

But to deny that irrelevant drives in the conditioning situation are responsible for the observed differences in the rate of conditioning between, say, introverted and extraverted subjects does not deny the importance of _relevant_ drives in both simple and complex learning situations. Degrees of hunger affected the conditioned salivation of Pavlov's dogs, but did not acfect the conditioned eye-blink (13). It was relevant in the former, irrelevant in the latter. But stress, and the resultant anxiety or neuroticism, could be regarded as a very relevant drive regarding a complex process like learning (or,

indeed, teaching). There has been much success, and much agreement, in demonstrating this to be the case. Spence and his associates predicted that in a situation where several response tendencies, including several incorrect response tendencies, were possible (i.e. in a complex task), the presence of high anxiety in a subject should make for poorer performance than low anxiety. This prediction has been borne out in a whole series of projects (1, 2, 6, 7, 12, 14, 15, 16, 17, 18, 19, 20, 22, 23, 31).

A valuable statement regarding the effect of drive on learning and performance was provided by the so-called Yerkes-Dodson law (Yerkes and Dodson, 1908), a law that has been familiar to psychologists for over fifty years. They found that the relation between motivation or drive on the one hand, and learning or effective performance on the other was curvilinear. The optimum drive level usually lay below the highest level induced in the experiment and above the lowest. Thus the intermediate levels of induced drive produced the best performance. The level at which drive was most effective depended, too, upon the complexity of the task: the more complex was the task the lower was the optimal drive level; the simpler the task, the higher the drive level. Young (32) reports that the Yerkes-Dodson findings have been confirmed for both animal and human subjects. In fact, probably most people would intuitively agree that it is the case in their own experience: a very strong drive like anger or fear is excellent for making a person perform a simple response like running or fighting, but a rather

inefficient motivator for carrying out a complex intellectual or manipulatory task.

Contribution to a Solution of the Problem

The foregoing studies are pertinent to the main investigation in this thesis in that they deal with anxiety and task performance. In these studies the concept of drive was central. Sometimes manipulation of the criterion (complexity of task) was attempted, sometimes it was manipulation of the independent variables level of anxiety or neuroticism, or the stress stimulus). In the problem of the present study, it is already assumed that the subjects (persons teaching in a classroom) vary in their individual levels of anxiety or neuroticism, and that they are subjected to a variety of situations that may be perceived as stressful and, as in certain situations of individual supervision, occasionally the stress stimulus is intensified. The complexity of the task has been assumed as constant. Although teachers quite frequently make comparisons as to who has the heavier task (me or the one across the hall), the essential nature of teaching requires that the task is everywhere at a reasonably high level of complexity. Thus there are three important factors here; one of which is assumed constant: (1) individual differences in the personality of the subject, particularly in level of anxiety, or neuroticism, or other characteristics that will tend to cause the subject to perceive a situation as ego-threatening (stressful), (2) the stress potentiality of the situation, i.e. drive stimulus, and (3) complexity of the task.

In somewhat different terms, but clearly related to the above factors, Eysenck has expressed a relationship which, coupled with the Yerkes-Dodson law, hypothesizes a solution to the basic question of this study:

> In addition to drive and response complexity as determiners of performance, we must also introduce individual differences in what might perhaps be called 'arousibility' or emotionality; thus the emotional anxious or neurotic (sic) person might even react to a very slight electric shock in a manner suggesting that for him this is a very strong drive, whereas the more stable, non-emotional kind of person might be able to accommodate quite a strong shock without allowing it to interfere with his efficient learning and performance habits. To the two dimensions of 'strength of drive' and 'difficulty of problem' contained in the Yerkes-Didson formulation, we must therefore add the third dimension of drive arousibility or emotionality; this dimension would be predicted to act in the same way as that of strength of drive. (11, p. 98)

The evidence in favour of this hypothesis is somewhat indirect. Eysenck has demonstrated that neurotics are inferior to normals, matched for age, sex, and intelligence, in their performance on complex motor tasks (8, 9). Broadhurst (3, 4) has conducted an interesting investigation under this three variable hypothesis using rats. He selected his animals as emotional or non-emotional on the basis of a selective breeding experiment and a defecation-urination test that was developed by Hall in 1934 and has since been reported valid for this purpose. They were then submitted to varying degrees of induced drive (the stress situation) by emersing them underwater for varying degrees of time. They then had to swim underwater through a Y-maze to get air. The maze indicated the

complexity of the task since they had to discriminate between the intensities of two lights to learn the correct way out, the light-ratios being varied to make discrimination more or less difficult. The experiment produced results to indicate the validity of Eysenck's hypothesis and the Yerkes-Dodson law. In our own problem, however, it has been assumed that persons show individual differences not only in arousibility level (neuroticism, anxiety) but also in the nature of their fundamental self-actualization needs and self-expectations. What may be a stress situation for one may not be for another with the same level of neuroticism, depending on the peculiar self-structures of the respective persons and the relevancy of the stress stimulus to them.

An important contribution of the discussion in this chapter is that behavior does not necessarily deteriorate under conditions of increased conflict (or increased stress, or increased drive stimuli). Performance under stress should, according to the Yerkes-Dodson law, go either up or down depending upon individual differences in personality. More specifically, under the Eysenck hypothesis one would predict interpersonal stress to produce a decrement in performance if it tended to push the individual past the point of optimum drive, but an increment in performance if it brought him up to the point of optimum drive. Fuller development of this concept appears in the next chapter.

REFERENCES FOR CHAPTER VII

(1) Axelrod, H. S., G. L. Cowen and F. Heilizer "The Correlates of Manifest Anxiety in Stylus Maze Learning." Journal of Experimental Psychology, 51 (1956) 131-138.

(2) Beam, J. C. "Serial Learning and Conditioning Under Real Life Stress," Journal of Abnormal and Social Psychology, 51 (1955) 543-551.

(3) Broadhurst, P. L. "Determinants of Emotionality in the Rat, I: Situational Factors," British Journal of Psychology, 48 (1957) 1-12.

(4) Broadhurst, P. L. "Determinants of Emotionality in the Rat, II: Antecedent Factors," British Journal of Psychology 49 (1958) 12-20.

(5) Buchwald, A. E. and H. G. Yamaguchi "The Effect of Change in Drive Level on Habit Reversal," Journal of Experimental Psychology, 50 (1955) 265-268.

(6) Castadeda, A. and D. S. Palermo "Psychomotor Performance as a Function of Amount of Training and Stress," Journal of Experimental Psychology, 50 (1955) 175-179.

(7) Deese, J., R. S. Lazarus and J. Keenan "Anxiety, Anxiety Reduction and Stress in Learning, "Journal of Experimental Psychology, 46 (July 1953) 55-60.

(8) Eysenck, H. J. Dimensions of Personality. London: Routledge and Kegan Paul, 1947.

(9) Eysenck, H. J. The Scientific Study of Personality. London: Routledge and Kegan Paul, 1952.

(10) Eysenck, H. J. "A Dynamic Theory of Anxiety and Hysteria," Journal of Mental Science, 101 (Jan. 1955) 28-51.

(11) Eysenck, H. J. The Dynamics of Anxiety and Hysteria. London: Routledge and Kegan Paul, 1957.

(12) Farber, I. E. and K. W. Spence "Complex Learning and Conditioning as a Function of Anxiety," Journal of Experimental Psychology, 45 (1953) 120-125.

(13) Franks, C. M. "Personality Factors and Rate of Conditioning," British Journal of Psychology, 48 (1957) 119-126.

(14) Gaier, E. L. "Selected Personality Variables and the Learning Process," Psychological Monographs, 66:17 (1952) Whole No. 349, Pp. 28.

(15) Gordon, W. M. and D. E. Berlyne "Drive-Level and Flexibility in Paired Associate Nonsense-Syllable Learning," Quarterly Journal of Experimental Psychology, 6 (1954) 181-185.

(16) Heilizer, F., H. S. Azelrod and E. L. Cowen "The Correlates of Manifest Anxiety in Paired Associates Learning," Journal of Personality, 24 (1956) 263-474.

(17) Hughes, J. B., J. L. Sprague and A. W. Bendig "Anxiety Level, Response Alteration, and Performance in Serial Learning," Journal of Psychology, 38 (1954) 421-426.

(18) Lazarus, R. S., J. Deese, and R. Hamilton "Anxiety and Stress in Learning: The Role of Intraserial Duplication," Journal of Experimental Psychology, 47 (1954) 111-114.

(19) Lucas, J. "The Interaction Effects of Anxiety, Failure and Intraserial Duplication." American Journal of Psychology, 65 (1952) 166-173.

(20) Montague, E. G. "The Role of Anxiety in Serial Rote Learning," Journal of Experimental Psychology, 45 (1953) 91-96.

(21) Mowrer, O. H. "A Stimulus-Response Analysis of Anxiety and its Role as a Reinforcing Agent," Psychological Review, 46 (1939) 553-565.

(22) Osler, S. F. "Intellectual Performance as a Function of Two Types of Psychological Stress," Journal of Experimental Psychology, 47 (1954) 115-121.

(23) Ramond, C. K. "Anxiety and Tasks as Determiners of Verbal Performance," Journal of Experimental Psychology, 46 (1953) 120-124.

(24) Selye, Hans. The Story of the Adaptation Syndrome. Montreal: Acta, Inc., 1952.

(25) Spence, K. W. "Learning and Performance in Eyelid Conditioning as a Function of Intensity of the UCS," Journal of Experimental Psychology, 45 (1953) 57-63.

(26) Spence, K. W. and I. E. Farber "The Relation of Anxiety to Differential Eyelid Conditioning," Journal of Experimental Psychology, 47 (1954) 127-134.

(27) Spence, K. W., I. E. Farber and J. A. Taylor "The Relation of Electric Shock and Anxiety to Level of Performance in Eyelid Conditioning," Journal of Experimental Psychology, 48 (1954) 404-408.

(28) Spence, K. W. and J. A. Taylor "Anxiety and Strength of the UCS as Determiners of the Amount of Eyelid Conditioning," Journal of Experimental Psychology, 42 (1951) 183-188.

(29) Taylor, J. A. "The Relationship of Anxiety to the Conditioned Eyelid Response," Journal of Experimental Psychology, 41 (1951) 81-82.

(30) Taylor, J. A. "Drive Theory and Manifest Anxiety," Psychological Bulletin, 53 (1956) 303-320.

(31) Taylor, J. A. and K. W. Spence "The Relationship of Anxiety Level to Performance in Serial Learning," Journal of Experimental Psychology, 44 (1952) 61-64.

(32) Young, P. T. Motivation and Behavior. New York: Wiley and Sons, 1936.

CHAPTER VIII

THE INTERPERSONAL DYNAMICS OF SUPERVISORY RELATIONSHIPS: A THEORETICAL ORIENTATION[1]

To answer the problem of this thesis it will be necessary first to integrate certain concepts of administration, personality, and learning into a single theory, and to investigate the validity of the theory by testing the hypotheses which it suggests. The necessity of theoretical preoccupation is most aptly justified with an aphorism attributed to Kant:

> Experiment without theory is blind;
> Theory without experiment is lame.

To begin with, the theory would assume -- in agreement with Sullivan's interpersonal theory as well as the existentialist writings of Jean-Paul Sartre (4) -- that most interpersonal situations carry stress potency, and that situations in which a superior-subordinate evaluational relationship is perceived or accepted (whether formally defined or only tacitly, even unconsciously accepted) have a greater possibility of actualizing this potency, at least in the subordinate. Where ego-threat is present, so is interpersonal stress.

Having thus assumed, the theory would state that a

---

[1] For the essential antecedents of this chapter, the reader is referred to Chapters Two, Five and Seven.

subordinate placed in an ego-threatening environment -- one which would generally be thought of as stress-inducing -- is not necessarily prevented from performing a cognitive task like teaching (a complex task involving the total personality of the teacher) at his optimum level of effectiveness but may -- depending chiefly upon both the cognitive and conative aspects of his personality -- have his performance level raised, lowered or remain unchanged.

Three Personality Factors

Whether or not the subordinate's effectiveness under conditions of stress-inducing supervision, relative to normal or "no-stress" conditions, improves, deteriorates or stays the same depends upon three specific aspects of his personality (and the degree to which he perceives the event as stress-bearing which, in turn, also depends upon these same aspects):

1. degree of neuroticism or anxiety,
2. level of scholastic aptitude of general intelligence,
3. strength of need for achievement.

Each of these is elaborated in the following paragraphs.

Neuroticism. The first factor may be described as proneness to neurotic breakdown under stress or as any other of the Eysenck synonyms for neuroticism such as anxiety, emotional lability, drive arousibility, perhaps even "frustratibility". This characteristic, presumed to be distributed on a scale from near-zero to near-infinity and present to some degree in most persons, would tend to determine the extent to

which an individual would develop a state of stress in an interpersonal event such as the one described.

Scholastic Aptitude or General Intelligence. The second factor refers to the individual's level of what is commonly termed scholastic aptitude or general intelligence. Measures of scholastic aptitude customarily reflect not only inate ability but also the product of the individual's past experience in coping with scholastic or cognitive tasks provided by the environment, that is his past involvement in evaluational interactions both personal and objective. In the language of Hebb (1), they tap intelligence of both "Type A" and "Type B". The two cannot be separated for measurement purposes but it is thought that scores on tests of the "scholastic aptitude" type, as compared with a more "culture-free" type (e.g. Raven's Matrices), are somewhat more the result of "B".

This variable should operate in the same way as neuroticism, although not for the same reasons. Instead, we invoke the personality theory of Carl Rogers and the interpersonal theory of Harry Stack Sullivan to show that past experience with evaluational interactions becomes assimilated into the concept of the self and that this is a determiner of individual behavior in future evaluational interactions. The self-concept, built up over years of an individual's personality development, represents his self-expectations of success or failure of performance on similar and related tasks. Because of the organism's fundamental need to maintain and enhance the self-concept, the

individual will tend to expect future performance-evaluations to be reasonably consistent with the past. Relating this to the problem of change in performance under various methods of supervision, one would state that when administrative expectations are contrued as being congruent with self-expectations, the former are readily assimilated into the self-structure -- assimilation is difficult and, unless the transmissal of external expectations is accompanied by a complete absence of ego-threat, a period of maladjustment follows. The experiencing of ego-threat is accompanied by emotional tension or psychological stress. Therefore, one should expect an individual who has a high degree of measured scholastic aptitude to experience psychological stress when placed in a situation where a significant authority-figure reveals an evaluation indicating cognitive failure. This would be so whether the authority-figure used the word "failure" or simply revealed administrative expectations to be so high that the individual cannot help but deduce that she had failed. It would, indeed, be so in any instance where perceived ego-threat was perceived by the subordinate.

The effects of this variable, of course, do not determine all of behavior. Perception of ego-threat depends, as well, upon the other variables listed in this chapter and possibly, others. Scholastic aptitude is but one of them.

Need for Achievement. The third factor reflects a desire to do a good job, to achieve success in endeavours of the sort under question. Generally, this has been described

as the need for achievement, or 'the achievant attitude'. Unfortunately, despite exhaustive literature in this field (e.g. 3) there remains some confusion as to its application to specific problems. The need for achievement seems to be compounding of several factors. Gough (2) believes there to be two such needs, to achieve via conforming and to achieve via exercizing independence. There is the added difficulty of sorting out genuine motivation to do a good job from the desire to be recognized as achieving success. But perhaps even more basic to the present problem is the question of what constitutes an unmet need, or frustration. Is it an awareness that a strong need has not been realized, or the realization that the environment actively inhibits need-fulfilment? In line with the Getzels suggestion that need-role conflict occurs when personal needs are perceived as incongruent with organizational expectations, one would expect conflict to develop in either of two quite different cases of need-frustration: (a) the teacher who knows how to teach first grade reading and who strongly desires to do a good job of it but who is so circumscribed by her overdirective supervisor's expectations that the guidebooks, workbooks, grouping systems and rates of progress be meticulously followed that she finds herself prevented from doing the job she wants to do, or (b) the teacher who has a strong desire to teach successfully and in accordance with supervisory directives but who finds that the supervisor's expectations for her performance are beyond that which she has accomplished -- her best was not good enough.

There is little doubt that both situations illustrate a frustrated need for achievement. Perhaps the confusion lies not so much in the personality attribute as in the type of stress engendered. Referring back to the Lazarus classification of stress types, situation (a) would be referred to as cognitive-interference stress relating particularly to the rationality of the task; situation (b) is a clear example of ego-threat failure stress. Both represent need-role conflict and, according to Getzels' model, should result in changed level of performance.

In each of the two examples, however, a strong need for some kind of achievement was hypothesized. What of a person with a weak need for achievement? In Chapter Seven it was noted that research reported by Eysenck indicated that non-relevant stimuli had little or no effect on the reaction in question. The relevance of the stimulus to the need or drive appeared as an important factor. If one ignores, for the moment, the contribution of factors of neuroticism and self-expectations, a prediction would be that frustration and conflict will develop in proportion to the relevance of the "frustrating situation" to strong personal needs. This agrees with Getzels' needs-expectations dimension in its effects. Perceived failure or threat of failure would logically cause greater frustration in a person possessing a strong need for achievement than in one who lacked this need. Change in effectiveness will, accordingly, depend upon whether the need for achievement is strong (present) or weak (relatively absent). For degree and direction of change in effectiveness, one is referred to the

following sections.

Directions of Performance Change under Stress

In general terms, the effect of psychological stress on task performance should tend to follow a curve. That was concluded in Chapter Seven. The relationship should operate approximately as follows: assuming task complexity to be constant, the stress stimuli combine with personal tendencies within the individual to produce a drive level which, if the stress increment had raised the drive level from below the optimum to the optimum, would result in a performance increment but which, if the stress increment had raised the drive level from the optimum to above the optimum, would result in a performance decrement.

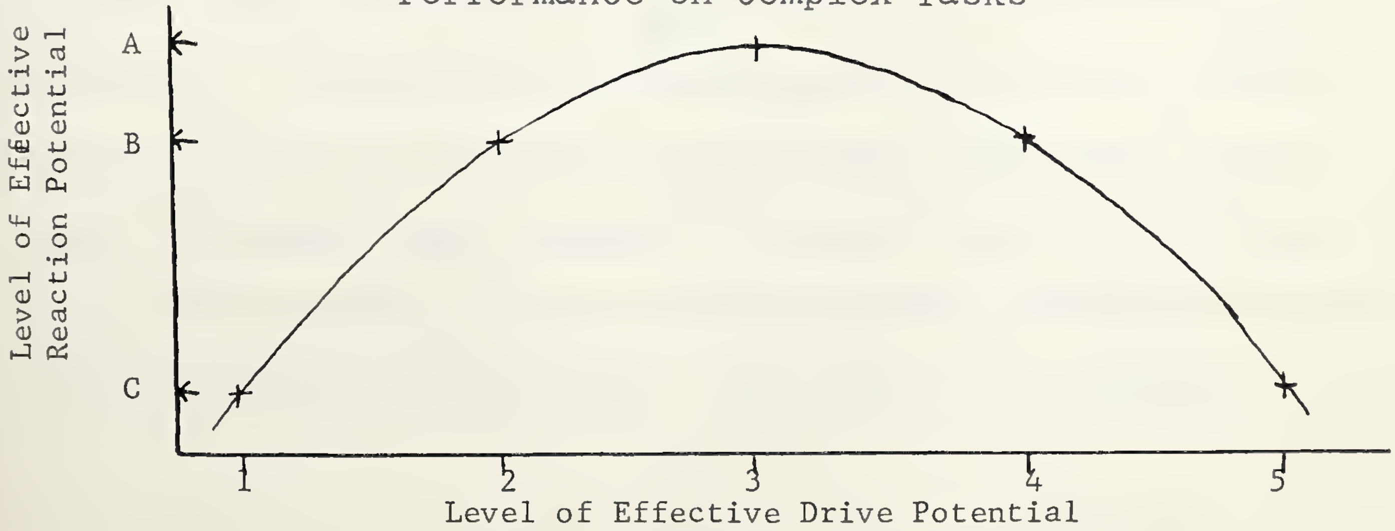
The three personality factors just discussed (and possibly others) combine to form what may be called drive arousibility, a reasonably constant personality attribute within the individual but varying from person to person from a low intensity to a high. There is a certain amount of stress present in every task (this is certainly true of teaching) which may be referred to as the normal challenges of the job -- this is part of the objective drive stimulus or factors within the environment that are reasonably constant for all individuals within a given class or activity. The objective drive stimulus is increased if, during individual supervision, a situation is created which could normally be regarded as stressful. This increment in objective drive stimulus one may experimentally

hold constant *in its presentation* (not, of course, in its effects: that depends upon drive arousibility factors). These two classes of factors--the internal and the external, the individual and the institution, the personality and the role-expectations, the drive arousibility factors and the objective drive stimuli -- combine (whether summatively or multiplicatively does not matter for this treatment) to form the total *effective drive potential* which varies (as a result of its component variables) from low to high.

Effective reaction potential means the individual's potential for producing behavior that is effective relative to the goals set for the complex task in question, in our instance: teaching. This also ranges from low to high. It is a dependant variable and, holding constant the variation from person to person in their specific abilities for accomplishing this task, varies curvilinearly with the total drive potential. When drive is low, effective performance is unlikely. Effective performance is optimum when drive level falls somewhere between the two.

Figure 4

Curvilinear Relationship of Drive Level to Effective Performance on Complex Tasks



where: Effective Reaction Potential depends upon Effective Drive Potential

and: Effective Drive Potential is the combination

of: 1. Drive Arousibility Factors
(a) neuroticism/anxiety; (b) scholastic aptitude; (c) need for achievement; (d) possibly others,

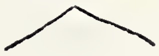
with: 2. Objective Drive Stimuli:
(a) normal challenge and frustration of the task, including its complexity; (b) any other extrenally-induced state, such as by stress-inducing supervision.

In Figure 4, the performance of an individual whose level of effective drive potential was optimum for the task, i.e. fell approximately at point 3, would deteriorate if his drive level were increased, i.e. moved toward 4 or 5, because his position of the curve would move downward to the right. An individual whose drive level was already somewhere about #4 would, of course, deteriorate even more. Conversely, it is possible that an individual with low drive, say at points 1 or 2, would, under conditions that increase overall drive, improve his performance by virture of his moving farther to the right on the curve and, therefore, higher towards the optimum.

Interpretation of Figure 4, as of this theory, requires that one think of a curve of this nature for each individual. A person with absolutely no ability for working at a task (say, an excellent lawyer given that task of performing delicate brain surgery for which he has no training or aptitude) would not start to perform with the best of them by increasing his drive level. The curve is of effective

reaction potential, not of positions on a scale of absolute effectiveness, and is therefore relative to the individual. The optimum point, accordingly, does not mean "top teaching" but the best teaching that one could expect this individual to perform.

Degree of Performance Change under Stress

Speculations on the degree of change that will occur under conditions of increased stress imply a statement as to the actual shape of this curbe. Unfortunately, there is no guidance readily available in this regard. If it were not actually a curve but an angle (e.g.: ), then one would conclude that differences were in direction only and not in degree. Moving from point 1 to point 2 would result in the same degree of change in reaction potential as from any point on the base to the next higher point; only direction would change. But this does not appear to be consistent with general observation or "common sense". Under constant conditions of stress it would seem that a person who is already quite neurotic to begin with develops reactions that are far less effective in coping with the situation than does one who is only moderately so. Their loss, if it could be quantified, seems to be greated than a person with only normal neurotic tendencies. This could be represented diagrammatically by an arc of an ordinary circle: the baseline or secant representing (as in Fig. 4) drive levels divided into equal intervals. Thus an equal increment in drive level would be shown as resulting in different degrees of increment or decrement in reaction

potentials. Ultimately, if the secant became the diameter, there would be a sharp falling-off at the upper extreme and a sharp gain at the lower. The same could, of course, be represented by a parabolic curve and would tend to accentuate these effects.

One further speculation: stretching the outer extremes of drive level far enough begins to suggest very abnormal circumstances. At the outer extremes one would possibly find persons institutionalized with severe neuroses, on one hand, or as a veritable 'vegetable' on the other. The point is that, because persons at very extreme ends of the drive scale in the total population would likely be rather ineffective (absolute terms) anyway, an increase in drive level would probably not have a great effect upon this effectiveness level. This is only speculation; it implies that for the total population the curve would flatten out toward the ends. That this begins to resemble the "normal probability curve" has no more than coincidental significance. Since the present study is concerned only with normals, however, speculation as to the abnormal extremes will remain just that. Figure 4, then, which was "just a curve", implies all that is necessary to illustrate the degree of change hypothesized among normals. The numerical values of the figure are arbitrary and serve only to represent in picture form the notion that equal increments along the baseline result in unequal increments along the ordinate.

Mean Effect

To suppose that the effect of an increase in stress does produce a change in effectiveness of performance depending upon individual differences is one valuable contribution. But administrators are concerned with overall tendencies as well as with tendencies due to individual personalities. Despite the central reference to individual supervision in this thesis, it must be recognized that much of the supervisor's time is taken up with group activities and planning; even when it is directed toward the individual the knowledge of some general or probable effect would be helpful. The theory can be of assistance here, too, if one assumption is allowed: that in any large number of persons working under reasonably good conditions, their respective drive levels will tend to be bunched up about some average point falling approximately midway between the two extremes. If not optimum, it may be thought of as nearly so for many of the subjects. Although it is sometimes said that the average teacher works far below her capacity it is doubtful that, all things considered, she is that far below. Teacher-trainees, under close supervision, would probably be near-optimum.

Granting this assumption it would follow that the introduction of a very powerful stress agent such as stress-inducing supervision, would tend to raise drive level past the optimum point and onward to produce some decrement in performance. In view of the speculation that differences in effect are not so great near the middle, and the notion that the

average person may be a little less than optimally motivated to start with, the net decremental effect would not be great. Thus, out of a large group one would expect the same stimulus to result in an average effect of a slight decrease in the effectiveness of the individuals taken as a whole.

Independent Variation

The theoretical development so far has begged the question of independent variation among component factors of personality, those contributing to the overall drive arousibility within the subject. This is not answered by the theory and assumptions about it are difficult to propose. True, several personality assessment instruments can demonstrate low intercorrelations but even if those three listed in this chapter are correlationally independent, or nearly so, to predict the outcome of an experiment where some subjects have very low "scores" on one factor, and high on another is problematical.

Eysenck suggested that the external and internal variables combine to produce total drive state. Taylor suggested that drives or needs summate to form effective drive potential. Whether the "internal" factors combine summatively, multiplicatively, or according to some kind of weighting, it seems clear that their eventual combination is accepted. The simplest solution is to propose that they be added or averaged.

Implications for Supervision

If the theory is to be at all generalizable, it will be necessary to pause over the nature of the experimental variable, termed externally-induced psychological stress, which

makes up part of the objective drive stimulus. Elsewhere this was referred to as any kind of superior-subordinate conflict situation, or lack of congruence and harmony between a teacher's needs and his perceived expectations. For administrators, the value of the theory lies not so much in what happens to classroom teaching when failure-stress is deliberately induced, nor is it limited to what happens when such stress is inadvertently aroused. Rather, it lies in what may happen to teaching effectiveness when a part of the objective drive stimulus is increased. This, after all, is a major function of supervision. There is little the supervisor can do about the antecedent factors of drive-arousibility or the teacher's personality. But his opportunities for manipulating drive stimuli are legion. For illustrative and experimental purposes the focus of this study has been on one kind of manipulation only: inducing the kinds of tensions that are aroused when one's ego is in danger. This was called stress. But to Selye, stress is something that accompanies virtually all our waking endeavours; he once defined it simply as the rate of wear and tear on the body at any one time (5). There is just a good deal more wear and tear in being bawled out by the boss than in listening to a symphony.

To persons occupying supervisory roles the point is that since any event, program or activity designed to stimulate and challenge teachers will have the effect of increasing the objective drive stimulus for each teacher (or is expected to), such activities should be carefully guided so that the result is

a motivation to react at the optimum level warranted by their teaching abilities. Beyond this, a teaching decrement could be expected. And, for most teachers, the probabilities of reaching this supra-optimum level are enhanced under conditions of severe stress, particularly including ego-threatening stress. Reaching only the infra-optimum level would be likely in a program devoid of interpersonal quality. Similarly, increasing drive level, it was earlier noted, does not produce an absolutely high level of teacher effectiveness -- simply a higher or lower level of performance in relation to the teacher's own potentialities. That is, in-service education of a motivational nature does not substitute for pre-service education. Increasing the drive level of a poorly-qualified teacher could easily make her more aware of her own inadequacies, resulting in marked personal stress and decreased teaching effectiveness. She may, of course, decide to obtain better pre-service education which probably is a decision that could better be arrived at directly through teacher-consultation. It is equally significant for the supervisor, in individual supervision situations, to recognize the necessity of differentiating motivational stimuli according to individual differences found among his teachers. This implies: that he must somehow become sufficiently acquainted with his staff to be able to differentiate.

Although the preceding paragraph refers chiefly to the motivational function of supervision, the theory has come implication for the consultative function as well. In so far

as the "normal challenge of the job" was included as part of the objective drive stimuli, it would follow that any activity or consultation that had the effect of increasing the teacher's awareness of the challenges of the job would result in the perception of discrepancies between what the teacher had thought the job to be and what she now realizes it is. This could be interpreted as mild conflict, some stress increment, and -- for the person who had not so realized before but went on operating at level below the optimal -- probably some increased drive. This is stimulation in the desired direction, toward learning more so as to increase her competencies to meet the demands of the task.

REFERENCES FOR CHAPTER VIII

(1) Hebb, D. O. A Textbook of Psychology. Philadelphia: W. B. Saunders and Co., 1959.

(2) Gough, Harrison B. The California Psychological Inventory and Manual. Palo Alto: Consulting Psychologists Press, 1956.

(3) McClelland, D. C., J. W. Atkinson, R. A. Clark and E. L. Lowell. The Achievement Motive. New York: Appleton-Century Crofts, 1953.

(4) Sartre, Jean Paul. L'Etre et le Neant (H. E. Barnes tr.: Being and Nothingness). New York: Philosophical Library, 1956.

(5) Selye, Hans. The Story of the Adaptation Syndrome. Montreal: Acta Inc., 1952.

CHAPTER IX

BASIC HYPOTHESES

The basic hypotheses of this thesis are those which are generated directly by the theoretical development in Chapter Eight; such was the purpose of that chapter. "The function of theory," Hebb believes, "is to guide us to new observations and better experiments" (1, p. 260). More epigrammatically, he notes that this function is better seen if we regard theory as "a sophisticated statement of ignorance" (1, p. 259). 'Sophisticiated' because the ignorance is put in a form that shows us how to go about decreasing it. The enunciation of the hypotheses follows the basic hypothetical solutions provided by the theory. Thus, it is hypothesized:

Hypothesis One

That teachers placed under conditions of stress-inducing supervision will, relative to control conditions, vary both in degree and direction of change in teaching effectiveness; some will deterioriate to different degrees, some will improve to different degrees, and some will remain the same.

Hypothesis Two

That the distribution of differences (a difference is the magnitude of change in effectiveness from normal conditions to stress conditions) will approach that of a normal distribution.

(hypotheses, continued. .)

Hypothesis Three

That the general or average tendency of teachers under conditions of supervisory tension will be in the direction of decreased teaching effectiveness.

Hypothesis Four

That increasing the external stress stimulus through ego-threatening or failure stress will depress the effectiveness of high-neuroticism subjects and will enhance the effectiveness of low neurocitism subjects.

Hypothesis Five

That increasing the external stress stimulus through ego-threatening or failure stress will depress the effectiveness of subjects evidencing experience of high scholastic aptitude and will enhance the effectiveness of subjects evidencing experience of low scholastic aptitude.

Hypothesis Six

That increasing the external stress stimulus through ego-threatening or failure stress will depress the effectiveness of subjects possessing a strong need for achievement and will enhance the effectiveness of subjects possessing a weak need for achievement.

(hypotheses, continued. ..)

Hypothesis Seven

That the three personality variables of Hypotheses 4, 5 and 6 vary independently, both among themselves and among their effects upon performance; that is their effects are additive.

Hypothesis Eight

That, taking Hypotheses 1 to 7 together, the direction of change in teaching effectiveness of subjects submitted to stress-inducing supervision is predictable from a combination of the three personality variables stated therein; the higher the combination, the greater the liklihood of a performance decrement, the lower the combination the greater the liklihood of a performance increment -- admitting, as an effect of the general or mean tendency, that those with a median combination will show a slight decrement and that the decrement among the high's will be greater than the increment among the low's.

Hypothesis Nine

That there will be a greater degree of performance change (incremental or decremental) shown by subjects who are at the upper and lower extremes of the personality variables than by subjects who are in the middle range.

Hypothesis Ten

That Hypotheses 1 to 9 inclusive, which apply to general effectiveness in the teaching task, will also apply to each of several component aspects of teaching effectiveness.

Called "basic hypotheses", these ten are presumed to incorporate all the main elements of the theory in verifiable form. Actual statistical verification will, however, require their restatement into a form that meets the requirements of the statistical model employed. Customarily this form is that of the null hypothesis; for example Hypothesis Three may be restated as follows: There is no significant different between[1] criterion measure means of three groups of subjects classified as high-, medium- and low-neuroticism personalities. Statements such as this will be enunciated as the analysis of results progresses.

---

[1] This breach of grammar is accepted by authorities (e.g. 2, p. 212) for the purpose of avoiding confusion with "among subjects within groups."

REFERENCES FOR CHAPTER IX

(1) Hebb, D. O. A Textbook of Psychology. Philadelphia: W. B. Saunders and Co., 1959.

(2) Walker, H. M. and J. Lev. Statistical Inference. New York: Henry Holt: 1953.

CHAPTER X

OPERATIONAL DEFINITIONS: INSTRUMENTATION

To test the hypotheses of Chapter Nine requires (1) redefining the variables of the hypotheses in terms of the tools used to measure or apply them -- the 'operational definitions', (2) developing a means of using these tools with appropriate subjects -- the experimental procedure, and (3) setting a method for analyzing the data so that inferences can be drawn in support or non-support of the hypotheses -- the statistical design.

Each of the variables has already received formal definition. An operational definition is one that shows how a variable is defined by its use in the specific operation. This is a cricial point in any investigation; if the means used for working with a variable are consistent with its formal definition, then conclusions can be drawn in formal and general terms. This is the ideal. Usually the operational definition must be somewhat more restricted in its scope than the formal definition, hence conclusions must be qualified accordingly. If there is some question of the validity of the means used, or if an instrument fails to satisfy the accepted meaning of a term for given conditions, then the investigation may be one that is perfectly sound but for some alternate hypothesis. In a sense, the validity of a piece of research is a function of the degree to which its operational and formal definitions

are consistent.

The following are definitions of the variables in terms of the means used to measure them or, as in the case of stress, to apply them.

Need for Achievement. Attempts to quantify man's desire to achieve success by using a paper-and-pencil test have proved it to be a quite difficult task. Most of the writing in the field (e.g. 17, 18) has reported the use of projective devices, notably the Thematic Apperception Test (TAT) and its descendents, which require conditions of administration sometimes difficult to achieve in group situations and demands scoring and interpretation skills that are not readily acquired. Simple "level of aspiration" statements have occasionally been used as an indicator of need for achievement but their validity is problematical (16).

Two relatively recent tests of general personality that are being widely used in research and clinical work contain achievement scales. These are the Edward Personal Preference Schedule (EPPS) (6) and the California Psychological Inventory (CPI) (12); copies of each appear in the appendix.

In general terms, the EPPS purports to measure fifteen relatively independent personality variables called manifest needs. Edwards credits the origin of the test items, of the fifteen variables and of the theoretical antecedents of the test to Henry A. Murray's need-theory (22) which also generated

the TAT. The test was designed primarily for research and student counselling purposes. The claim that the test is relatively free from contamination by the social desirability of the item content is supported with research evidence (5, 6).

The CPI, described as "the sane man's MMPI," does not appear to be set within any particular theory of personality, but rather provides a convenient classification of a subject's responses to a very large number of empirically-derived items.

The "need for achievement" variable of this thesis is defined as a score on one of the EPPS scales, "n. Achievement", (abbreviated n Ach), and on two CPI scales: achievement via conformity (Ac) and achievement via independence (Ai).

The reliability of the n Ach variable is reported (6) in two ways: (a) a coefficient of internal consistency (corrected by the Spearman-Brown formula) of .74 (N=1509) and (b) a coefficient of stability (one-week separation) also of .74 (N=89). The fact that the theoretical antecedents of the EPPS are securely-based in well-established personality psychology attests to its construct validity. Edwards reports several predictive validity studies wherein self-ratings and ratings by peers, using simple rankings and Q-sorts, established a certain degree of predictive validity. Concurrent validity is inferred from correlations with several other personality measures which, Edwards reports, were in the expected directions.

Gough (12) reports no internal consistency reliabilities but produces test-retest coefficients of stability (seven to twenty-one days separation) for the Ac scale of .73 for male, .60 for females; for the Ai scale of .57 for males and .63 for females. Because of the lack of personality theory behind the CPI's development, construct validity is difficult to infer. Correlations with other personality variables are not reported, hence concurrent validity is similarly equivocal. Predictive validity is inferred by Gough from the following quite low but statistically significant correlations: Ac with highschool grades is .41; with intelligence quotients is .13 for females and .19 for males; Ai with Q-sorts referring to self-reliance and independence is .30; with university average marks is .44, and with grades in a psychology course is .38.

One must conclude that measures of the need for achievement leave some doubts as to their reliability and validity. Because there are no other measures available, findings based upon these measures must be regarded as little more than tentative.

Neuroticism and Anxiety. Attempts to construct an anxiety index began with variations on widely used existing tests. Of these, the Rorschach, TAT and Wechsler-Bellevue resulted in indifferent success (11, 14, 21, 25, 26). The Minnesota Multaphasic Personality Inventory (MMPI), probably the most widely known paper-and-pencil clinical instrument, has been used as a basic for no fewer than five different

anxiety indices, and all have been used with some success (32). Most of them employ a combination of existing MMPI scales as an index of anxiety; for example the anxiety index developed by Welsh (30) is a unique formula involving Hs, Hy, D and Pt scales of MMPI. The scale devised by Taylor (28) has, however, been the most successful and most generally accepted of them all, at least on this continent. Taylor's scale was built according to good test construction tradition using as its "pool" all MMPI items judged by clinicians as having relevance to anxiety. The pool was reduced through application and statistical means to a list of 50 items which were embedded in 175 filler items. Norms were based on administrations to 1971 students over a period of four years. Taylor gave the scale the innocuous title of Biographical Inventory; it has since been commonly referred to as the MAS (manifest anxiety scale) or the Taylor A-Scale. McCreary and Bendig (19) obtained a correlation of .95 between the complete scale and the 50 scored items administered separately and promptly questionned the necessity of the filler items. Although there have been some criticisms of the precision with which the A-scale measures manifest anxiety (10, 13), of its non-unitary nature (7), and of its validity in terms of the physiological aspects of anxiety (15, 24), the scale appears to have stood up well with time (3, 20, 23, 31, 32). A 28-item version of the A-scale was early developed and has frequently been used in research, including the TEASEL research project in teacher selection at the University ofAlberta. Numerous studies have found that the reliability and validity

of the A-scale are adequate for research and clinical purposes (reliabilities run from .70's to .90's, validity with several criteria run in the .50's to .90's excepting these criteria like palmar perspiration where correlations are quite low).

Another instrument that seems to measure the same general characteristic and which is particularly closely related to the variable under investigation in this thesis is the Maudsley Personality Inventory (MPI) developed by Eysenck (first reference 7, published test 9) for clinical use and for research in the fields of personality and learning. It contains two scales presumed, by Eysenck, to be the basic personality dimensions in the conative area: neuroticism and extraversion-introversion. The neuroticism scale purports to measure neurotic tendencies, or "the general emotional lability of a person, his emotional overresponsiveness, and his liability to neurotic breakdown under stress" (9). Regarding the confusion over the terms neuroticism and anxiety, Eysenck writes:

> This term (anxiety) often implies not only what is psychiatrically known as anxiety state, but the whole group of neurotic disorders, so that 'anxiety' and 'neuroticism' are used interchangeably. (7, p. 332)

The scales were derived by means of an extensive item-analysis from Guilford's STDCR and GAMIN questionnaires, and the Maudsley Medical Questionnaire; a correlation of -.05 (N = 400) has been found between the two scales (7, p. 329). A short-form, consisting of one-fourth of the MPI items correlates approximately .9 with the full scale (8). The reliability of

the test is indicated by both split-half and Kuder-Richardson coefficients on many samples; for Neuroticism they lie between .85 and .90, for Extraversion between .75 and .85. Retest reliabilities are .83 and .81 respectively. The validity of the test is indicated in a number of ways. Use of the test with deviant groups (N varies from 36 to 146) of known diagnosis turned up norms in the predicted directions, indicating the predictive validity of the instrument. An indication of the construct validity of the MPI, the validation given by its incorporation within a scientific system or nomological network, lies in the fact that the scales were built along with and as a result of Eysenck's hierarchical theory of personality. By way of concurrent validity the MPI manual (9) and other sources (4) cite studies showing that MPI-N correlates with the Heron and Cattell neuroticism scales, the Guilford C scale and the Taylor MAS (with MAS: .77), and provides similar correlates validating the MPI-E scale.

Data purporting to show the reliability and validity of measures in the affective or conative aspects of personality frequently leave many doubts lingering in the mind of the user. From those reported for the MPI by Eysenck and by several other writers (e.g. Bendig, Franks) one may safely conclude that this instrument is acceptable for the purpose of the present investigation. Although it is anticipating later findings in this study, the writer considers it psychometrically highly significant that a test-retest correlation between the MPI-N scores of 176 subjects in the testing sample, with testings

three and one-half months apart, was .75. Customarily test-retest coefficients are based on time-intervals of one to three weeks. An opportunity to re-rest our subjects three and one-half months after initial testing was taken in order to learn the stability of this personality variable over a long period of time. Although the obtained coefficient of .75 is lower than Eysenck's retest figure (.83), he does not report the time-interval and it may be assumed that it was brief. One may reasonably expect some personality change to take place over our longer period of time and, accordingly, a lower obtained correlation. Evidently one may conclude not only that Eysenck's test has high reliability but also that the neuroticism variable is one that is relatively stable in the 176 subjects tested.

Considering that the nature of the variable being investigated (see Chapters Eight and Nine) is the individual's susceptibility to stress reactions when placed under conditions of superior-subordinate conflict, it would appear advisable that a score on the neuroticism scale of Eysenck's test (abbreviated MPI-N) be used as the operational definition. A copy appears in the appendix. Although there is a substantial correlation between the MAS and the N-scale of the MPI, the fact that Taylor's anxiety scale has a significant loading on the extraversion factor -- which is not a part of the theory used in this investigation -- casts doubt upon its value here. In view of the ambiguity in the term 'anxiety' (see Chapter Six) the investigator decided to exclude the MAS rather than risk

confusing later interpretation of findings.

Scholastic Aptitude or General Intelligence. The choice of the American Council on Education Psychological Examination for College Freshmen (1, 29) (abbreviated ACE) was made primarily on administrative grounds. Test data was readily available on all subjects. This is the test that has been given to University of Alberta registrants for many years. That the manual (1) no longer publishes reliability and validity data is disconcerting. The latest *Mental Measurements Yearbook* indicates that although users find the test reliable, its validity varies from very high to questionable. The fact that it has received exceedingly wide use for over two decades indicates, perhaps, its general if not necessarily statistical acceptability. The test does not yield an intelligence quotient but scores the individual college means of which range from 131.03 to 45.36. The norms based on returns received from 293 colleges and 64,276 students show a mean of 105.24 with a standard deviation of 25.15. The practical significance of such a wide standard deviation is the greater degree of discrimination possible about the middle range.

Stress Situation. This is defined by procedure more than by instruments; the actual procedure will be described in an appropriate later section. Here it is important to note that the stress situation must be as controlled as possible since the personal factors entering into the overall drive level

are permitted to vary. Some mild stress will be present in virtually any activity; the experimental procedure must arrange for these "normal challenges and nuissances of the job" to be as similar, or controlled, as possible also.

To some subjects a given stressor may appear as very mild, to others it may be very strong. To reach all subjects in the sample, the stress agent must therefore not only be uniform but uniformly strong. This is possible, and in keeping with Getzels' theory, if during the process of individual supervision, the teacher perceives the institutional expectations for her role (as represented by verbal statements from the supervisor) as being incongruent with the teacher's total personality elements. Examples of such a generalized situation would be where (a) the teacher has a strong desire to achieve success in this teaching task but the supervisor acquaints her with the information that the institution expects far different results from that which she is achieving; (b) the teacher is rather bright and expects to be able to move smoothly through this task but is informed that, relative to organizational goals, her work is a personal failure; or (c) the teacher is so neurotic that she desires a smooth and ego-supporting relationship from her superiors but, instead, perceives their statements to contain a threat to her ego-security. Such examples are extreme both in implied teacher-personality characteristics and in supervisory activity; but for the purpose of uniformity of treatment the essential elements of these examples will be incorporated into a standardized

stress procedure.

Performance Change. Because the hypotheses do not predict absolute effectiveness but only the effectiveness of an individual's performance under stress relative to his effectiveness when the stress stimulus is absent, the criterion is a measure of change in performance. The hypotheses require that it be a measure both of direction and of degree of change.

There are two ways of arriving at such a criterion. One is to obtain two absolute measures of teacher-effectiveness on each subject, one measure for the control situation (Ec) and one for the experimental (Ee). The criterion measure would be calculated as the difference Ee - Ec. But the reliability of the measure will be affected by the number of abstrations required. In this method the experimenter would be required to make two independent real-to-hypothetical comparisons and to express each numerically as Ee and Ec. Inter-experimenter agreement becomes quite an issue with this procedure because of the possibility of different experimenters (i.e. raters) entertaining different concepts of the hypothetical or "ideal" standard. The absolute assessment of teacher effectiveness is an exceedingly complex process and only little headway towards objective measurement has yet been accomplished (2).

The second method achieves the same end result but with fewer abstractions: a simple comparison of the effectiveness of two "real" situations, the control and the experimental.

This does not relieve experimenters from the obligation to agree upon the "ideal" standard, particularly with respect to degree of change, but since the process is abstraction at a lower level the measure should be more reliable.

In this investigation the measuring instrument is entitled the "Confidential Report on Student Teaching (Research Adaptation of Forms 60-2, 50-2)" but became known during the project simply as the "White Rating Sheets". The format had been adapted to allow space for rating both the control and the experimental lesson on a single form to follow the "comparison" procedure, above. Faculty of Education Form 60-2 differs from 50-2 in the manner of rating the components only; the earlier one called for a 5-point rating but the later one left this open-ended.

There are a number of indications that the criterion variable measure used in this thesis possesses reliability. In a later section on procedure it is noted that the experimenters asked the subjects how they felt about the experiment and to make any comments they wished. These were written on the reverse of the White Rating Sheets by the experimenters. Unfortunately, they did not specifically ask each subject how successfully they thought they taught the second, as compared with the first lesson. In only about one-third of the instances did the subjects specifically indicate to the experimenters their own personal assessment of their behavior change from one lesson to the other. In about nine-tenths of these instances

the assessment of the subject coincided with that of the experimenter. This fact may be interpreted as indicating that two persons using the same criterion (change in effectiveness), one with a piece of paper and the other orally, assessed the variable in the same way.

A second, although minor, indicator came about even more informally. In each of six instances in which a regular faculty consultant gave an opinion to the acting faculty consultant (experimenter), after the experiment with the consultee in question, in which he 'predicted' how his consultee would have behaved under the stress conditions, it turned out to be just as the experimenter had rated the subject during the experiment.

A third bit of evidence deals not with the criterion of change in behavior under stress, but with the regular use of the Confidential Report on Student Teaching, Form 60-2, in the student teaching program. This form is used by both the regular faculty consultant and the cooperating teacher (these terms are explained fully in Chapter Eleven under "The Student Teaching Program") as a basic for evaluating regular student teaching. Each person submits a mark, the evaluations are arrived at independently and are based on different observations lessons during a round. The marks submitted by each are inspectionally so similar that a high correlation is suggested. Expressed ultimately in percentages, there were only a few subjects (of 182) for which there was a difference

of 15 marks, for none was it 20. For most the difference was 0 or 5 marks. Although this does not directly support the reliability of the "White Rating Sheet" it shows a time-to-time and user-to-user consistency in the instrument from which it was adapted.

Content and construct validity of the instrument receives a measure of support from the fact that the rating form for student teaching has been in use for eleven years (Form 50-2 for ten, 60-2 for one) at the Faculty of Education, University of Alberta, where the experiment was conducted. This inference is based on the principle that when both tester and tested are aware of the goals of the activity, test validity will be higher than when agreement is lacking. The subjects, the raters, and the personnel of the institution in general were familiar with the criteria of teaching effectiveness used in the student teaching program. If one were to rate the teaching performance of subjects who had previously been rated a number of times by members of the same institution, a truer evaluation of their teaching relative to institutional goals would be shown by the accepted rating instrument than on one which evaluated behavior relative to a different set of goals.

A part of this investigation was to determine whether, if interpersonal stress affected teaching activity at all, the effect would be more pronounced in some aspects of teaching than in others. The way a teacher uses his personality to

advantage with a class, his routine for maintaining an effective learning situation, and the extent to which he draws upon his preparation for the lesson each are aspects of teaching that may be differentially affected by supervision stress. The components of teaching included in the White Rating Sheet were those that were in the Forms 60-2 and 50-2: personal qualities, evidence of preparation for teaching, presentation of lession, contact with pupils, and classroom management. Copies of Form 60-2 of the Report on Student Teaching and the research adaptation appear in the Appendix. The research adaptation made it possible for the experimenter to signify that teaching performance under stress conditions, relative to control conditions, showed

- marked improvement, with a score of +3
- some improvement, with a score of +2
- slight improvement, with a score of +1
- no perceptible change at all, score 0
- slight deterioration, with a score -1
- some deterioration, with a score of -2
- marked deterioration, with a score of -3.

Thus, the criterion measure of performance change is defined as a "change-in-effectiveness" score, or D-score (difference-score) on a scale ranging from +3 through 0 to -3.

Classification of Variables

The variables defined in this chapter fall into the three broad classes suggested by the theory: personal or

antecedent variables, external or experimental variables, and the performance change or criterion variables. The theory was two-dimensional in that it construed the criterion as being a function of the combination of two classes of independent variables ('independent' in the sense that the criterion depends upon them and is thus dependent' not in the sense that they are independent of each other). In the theory these were termed the drive arousibility factors, and the factors of the objective drive stimuli. Relating these to Getzels' theory (setting aside the term "need-dispositions"), they become the personal or idiographic dimension including all teacher-personality variables, and the normative or nomothetic dimension including the institutional roles and expectations as perceived by the teacher. To Eysenck, these are related to his neuroticism or drive arousibility, and to the actual drive stimuli including the experimental stimulus and the complexity of the task. The variables defined in this chapter are classed as" antecedent or personal variables (n Ach, Ac, Ai, MPI-N and ACE), experimental or external variables (normal task complexity and challenges, and experimentally-induced stress). The dependent variables, or criterion, ('observed behavior' to Getzels, 'reaction' to Eysenck) is the D-score.

REFERENCES FOR CHAPTER X

(1) American Council on Education. *Manual of Instructions for the Psychological Examination for College Freshmen*. Princeton, N. J.: Cooperative Test Division, Educational Testing Service, 1950.

(2) Barr, A. S. D. E. Eustice and E. J. Noe "The Measurement and Prediction of Teacher Efficiency," *Review of Educational Research*, 25:3 (June 1955) 261-269.

(3) Bendig, A. W. "Age, Sex, and the Manifest Anxiety Test," *Journal of Consulting Psychology*, 18 (Feb. 1954) 16.

(4) Bendig, A. W. "Extraversion, Neuroticism, and Manifest Anxiety," Journal of Consulting Psychology, 21:5 (Oct. 1957) 398.

(5) Edwards, A. L. *The Social Desirability Variable in Personality Assessment and Research*. New York: The Dryden Press, 1957.

(6) Edwards, A. L. *Personal Preference Schedule* and *Manual* (Revised Edition). New York: Psychological Corporation, 1959.

(7) Eysenck, H. J. "Reminiscence, Drive, and Personality Theory," *Journal of Abnormal and Social Psychology*, 53 (1956) 328-333.

(8) Eysenck, H. J. "A Short Questionnaire for the Measurement of Two Dimensions of Personality," *Journal of Applied Psychology*, 42 (1958) 14-17.

(9) Eysenck, H. J. *The Maudley Personality Inventory and Manual*. London: University of London Press, 1959.

(10) Goodstein, L. D. "Interrelations among Several Measures of Anxiety and Hostility," *Journal of Consulting Psychology*, 18 (Feb. 1954) 34-39.

(11) Gorlow, L., C. N. Zimet and H. J. Fine "The Validity of Anxiety and Hostility Rorschach Content Scores among Adolescents," *Journal of Consulting Psychology*, 16 (Feb. 1952) 13-75.

(12) Gough, Harrison G. *The California Psychological Inventory*. Palo Alto: Consulting Psychologists Press, 1956.

(13) Kendall, E. "The Validity of Taylor's Manifest Anxiety Scale," *Journal of Consulting Psychology*, 18 (Dec. 1954) 429-432.

(14) Lindzey, G. and A. S. Newburg "Thematic Apperception Test: A Tentative Appraisal of some 'Signs' of Anxiety," Journal of Consulting Psychology, 18 (Dec. 1954) 389-395.

(15) Lotsof, E. J. and W. L. Downing "Two Measures of Anxiety," Journal of Consulting Psychology, 20 (June 1956) 170.

(16) McClelland, D. C. Personality. New York: Dryden Press, 1953.

(17) McClelland, D. C. Studies in Motivation. New York: Appleton-Century-Crofts, 1955.

(18) McClelland, D. C., J. W. Atkinson, R. A. Clark and E. L. Lowell. The Achievement Motive. New York: Appleton-Century-Crofts, 1953.

(19) McCreary, J. B. and A. W. Bendig "Comparisons of Two Forms of the Manifest Anxiety Scale," Journal of Consulting Psychology, 18 (June 1954) 206.

(20) Matarazzo, J. D. "MMPI Validity Scores as a Function of Increasing Levels of Anxiety," Journal of Consulting Psychology, 19:3 (June 1955) 213-217.

(21) Moldawsky, S. and P. C. Moldawsky "Digit Span as an Anxiety Indication," Journal of Consulting Psychology, 16 (April 1952) 115-118.

(22) Murray, H. A. Explorations in Personality. New York: Oxford University Press, 1938.

(23) O'Connor, J. P. and J. W. Stafford "Some Patterns of Manifest Anxiety," Journal of Clinical Psychology, 12:2 (April 1956) 160-163.

(24) Raphelson, A. C. "The Relationship among Imaginative, Direct Verbal, and Physiological Measures of Anxiety in an Achievement Situation," Journal of Abnormal and Social Psychology, 54 (1957) 13-18.

(25) Raphelson, A. C. and R. W. Moulton "The Relationship between Imaginative and Direct Verbal Measures of Test Anxiety Under Two Conditions of Uncertainty," Journal of Personality, 26:4 (Dec. 1958) 556-567.

(26) Stopol, M. S. "Rorschach Performance in Relation to Two Types of Stress," Journal of Consulting Psychology, 18 (1954) 11-15.

(27) Stopol, M. S. "The Consistency of Stress Tolerance," Journal of Personality, 23 (Sept. 1954) 13-29.

(28) Taylor, J. A. "A Personality Scale of Manifest Anxiety," Journal of Abnormal and Social Psychology, 48 (1953) 285-290.

(29) Thurstone, L. L. and T. G. Thurstone Psychological Examination for College Freshmen. New York: American Council on Education, Educational Testing Service, 1948.

(30) Welsh, G. S. "An Anxiety Index and Internalization Ratio for the MMPI, Journal of Consulting Psychology, 16 (Feb. 1952) 65-72.

(31) Wilensky, H. "A Factorial Study of Behavioral and Psychological Measures of Anxiety," Journal of Consulting Psychology, 21:3 (June 1957) 216.

(32) Windle, G. "The Relationships among Five MMPI 'Anxiety' Indices," Journal of Consulting Psychology, 19:1 (Feb. 1955) 61-63.

CHAPTER XI

EXPERIMENTAL PROCEDURE: SUBJECTS, PROGRAM AND EXPERIMENTERS

Five years ago a Canadian university professor, writing on the needs and dimensions of research in educational administration, urged researchers to question the basic assumptions used, to evaluate them and revise theory, to set up new hypotheses, and to test them scientifically by developing the controlled-experiment approach (1). The assumptions that were challenged in the present thesis were that need-dispositions exhaust the personality elements and that personal-institutional incongruencies necessarily had a deleterious effect on observed behavior; the revised theory called on personality theory and learning theory to show what additional personality variables must be accounted for and what would be the probably results of these in an institutional setting; the hypotheses are new in that they have probably never been advanced before and are not directly deducible from theories of administrative relationships; finally, a description of the experimental procedure and statistical design follows.

Briefly, the procedure was to:

(1) obtain a number of subjects who were engaged in the activity of classroom teaching,

(2) test them with standardized tests purporting to measure or imply presence of the antecedent variables,

(3) classify the subjects as placing high, medium, or low on each variable in order to predict their behavior change under stress,

(4) observe their teaching performance under control (no stress) conditions,

(5) submit them to stress-inducing supervision,

(6) observe their teaching performance under these stress conditions,

(7) signify degree and direction of change in teaching effectiveness (both globally and in terms of each of five components), such criterion measures to be called D-scores,

(8) extuinguish stress

(9) analyze the variance of the D-scores of the subjects classified on the bases of the antecedent variable scores, and

(10) test the hypotheses.

The Sample

The choice between using, as subjects, qualified teachers employed full-time in regular school systems (as the hypotheses may be thought to imply) or student teachers in training at the Faculty of Education, University of Alberta, was resolved in favour of the latter. Besides the presumed ease of obtaining a captive sample in this manner, there are several methodological reasons valid from the point of view of research design. Four of these are discussed.

(1) With a student-teacher population there is a better possibility of controlling a number of non-experimental variables: the educational and teaching backgrounds of the subjects, their age, the kinds of factors on which they are accustomed to being evaluated, and the standards of behavior

expected, to name but a few. The group could be selected in a manner that would make them relatively homogeneous on variables which otherwise could have a contaminating influence on the results.

(2) Within a university student-teaching program there is a better chance of obtaining the authority-dependency relationship necessary for a study in administrative relationships. In the field, this could not be obtained were the "authority-figure" or supervisor a stranger from without the system. Actual authority-figures within the system, if their cooperation as experimenters could be obtained, would vary greatly in the kind of administrative relationships they had established with their staffs over the years. Within a student-teacher program, however, the relationship established between students and their faculty consultants is much more uniform. The consultant, being a faculty member or graduate student, typically enjoys the "authority of knowledge" and a certain traditional status in relation to his consultees; in addition the student is dependent upon him with respect to student-teaching activity and his overall success within the institution. That is, faculty consultants award university marks for this program and frequently write recommendations to prospective employers. This does not mean that the entire teacher-superintendent relationship can be experimentally duplicated but that the essential elements required for a study of the effects of superior-subordinate stress upon the subordinate's effectiveness in a teaching activity will be present.

(3) Because testing on a number of antecedent variables was planned it is valuable for purposes of test interpretation that the sample be as similar as possible to that which the test-maker had in mind when the instrument was constructed and normed. The bulk of tests available (beyond those for use with schoolchildren) have been used with university students.

(4) The research cooperation of both the subjects and their superiors or authority-figures may more readily be maintained in a university setting than in the field. Hence, the validity of the results are enhanced. Both the students and the instructors at the Faculty of Education, University of Alberta, are considerably "research-oriented", possibly as a result of its being a university setting but more likely the result of the "conditioned acceptance" of research projects following from the sheer bulk of research carried on there over recent years. There is also a unified administration that is not present when a number of school units are used. As will be seen later, the choice of a location having potentially greater possibilities of cooperation and research orientation has very great importance to the validity and significance of the experiment and its results; this is not simply a matter of presumed ease of procedure.

Characteristics of the Sample

Initially, a sample of 182 subjects was drawn for the purpose of collecting data on the antecedent variables. From this number, 78 were included in the actual stress experiment.

The two groups (i.e. total group and sub-group) are referred to respectively as the testing sample and the experimental sample.

Of the testing sample (and, hence, of the experimental sample):

(a) all possessed clear senior matriculation from high school,

(b) all had completed one year of university, clear,

(c) all were registered in second-year university in the Faculty of Education,

(d) all were following the degree (B.Ed.) program offered by that Faculty,

(e) for most, this was their second year out of high school (it could not be less; for a few who had transferred to Education it was more),

(f) all were registered in the student-teaching program (course: Education 382) at the Faculty of Education, University of Alberta,

(g) all had completed two rounds (approximately two-thirds) of student-teaching during the current academic year prior to the time of the experiment.

From this it can be concluded that there was within the sample a high degree of homogeneity with respect to age, educational background and present program, entrance standard hurdled, program-identification (i.e., these persons customarily do not identify themselves with those on the one-year program for which most points on the above list would not apply), teaching experience (or lack of it), and exposure to faculty supervision in student-teaching. Clearly, if teaching under duress is affected differentially by any of the above variables, they have been controlled in this study.

The testing sample represented forty per cent. of the total number of students registered in second-year Education at October 1st, 1961.

Collecting Test Data

The subjects were tested, during regular classroom periods, on the MPI-N, n Ach, Ac and Ai scales during the weeks prior to the actual experiment. At this time they had no knowledge of the experiment nor of a connection of this testing with any future projects. While it is true that the validity of a personality measure sometimes suffers from the respondents' faking, it is generally believed that the validity is also a function of the rapport between tester and tested. The better the rapport, the less likely is the tendency to fake. In this instance, the tests were administered toward the end of the first term by which time a very good working relationship had been established between the tester (the university instructor) and the tested (his regular class).

Data for the ACE were made available from university records. These were approximately sixteen months old at the date of the experiment since the tests had been administered to the students when they first registered.

The Student-Teaching Program

Since the experiment was to be run within the regular channels of the student-teaching program, a brief description of it in so far as it relates to the experimental routine, is here provided. Students in the B.Ed. program enroll in

student teaching during their second university year, the year that qualifies them for a "standard" elementary or secondary teaching certificate. Their student teaching carries on throughout the year but is divided into a number of blocks of time, e.g. of two or three weeks duration, called "rounds". The elementary-route students, during a round, are in the city schools all day every day; the secondary-route students are in the schools during the afternoon only and return to the university for lectures each forenoon. The overall program is under the direction of the Supervisor of Student Teaching, aided by an Assistant Supervisor in charge of elementary-route students. Students are normally assigned to a single Cooperating Teacher (sometimes called a critic teacher) in a single school in the City of Edmonton for the duration of a round. The Cooperating Teacher assigns lessons to the student teacher.

There is also a member of the Faculty of Education assigned to each student teacher for the purpose of supervising, assisting and evaluating the classroom activity of the student. Called Faculty Consultants, these persons visit each of their student teachers at least once per round. A number of faculty consultants are graduate students working towards master of education degrees; they operate in the same way and with the same responsibility toward their consultees as do those who are full-time staff members of the Faculty. The Faculty Consultant evaluates the classroom teaching of his consultees using the "Confidential Report on Student Teaching,

Form 60-2" previously described. The student teacher is aware of this; aware, too, of the importance of this evaluation in terms of university standing and probabilities of success in teaching.

Clearance and Conditions

Permission to use undergraduate student teachers in this experiment was sought and obtained from a committee of appropriate persons at the Faculty of Education: the Dean, the Supervisor of Student Teaching, the Assistant Supervisor of Student Teaching, the Chairman of the Division of Elementary Education and the Chairman of the Division of Secondary Education.

Partly to assist in experimental design and partly to maintain good student and school relations, certain conditions were agreed upon:

(1) Consultants other than the students' regular faculty consultants must be used as experimenters.

(2) Cooperating Teachers must be acquainted in advance (in fact, not only were cooperating teachers but also regular faculty consultants and school principals acquainted in advance, in person).

(3) Induced stress must be extinguished effectively and immediately upon the close of each experimental visit.

(4) Students should have completed at least one round and,

preferably, two prior to being experimented with (two were completed).

(5) All persons involved in the project would be requested to maintain security.

The Experimenters

To carry out the experimental procedure successfully and within these conditions required experimental visits to all the selected subjects of one round during that round and not spread out over the year. Such concentration further required obtaining and training a number of experimenters who would be able to act the role of a faculty consultant.

It was earlier noted that a number of graduate students serve as regular faculty consultants. Several of these agreed to participate in this project as experimenters with students other than their own consultees. An additional number of graduate students, several of whom had regular teaching duties at the Faculty of Education, also served as experimenters. These persons, together with the present writer, totalling fifteen in number, were referred to as "Acting Faculty Consultants" during this project. Their characteristics and qualifications:

(1) All were intramural graduate students in the Division of Educational Administration, University of Alberta.

(2) Approximately one-half were M.Ed. candidates, one-half were first or second year Ph.D. candidates.

(3) All but three held regular duties teaching university classes or supervising student teaching.

(4) All had been in administrative or supervisory positions in education, had had experience in observing teachers in action and evaluating teaching effectiveness.

(5) Their educational experience ranged from 6 years upward.

(6) All were fellowship holders, having been selected from applicants across Canada for their ability and potential.

In order to prevent contamination of criterion scores none of the experimenters knew the position of any of the subjects on any of the test variables (this includes the present investigator, who arranged to have scoring done by another person). Permission to use these graduate students in this project was sought and obtained from the Chairman of the Division of Educational Administration. Permission for these persons to enter Edmonton schools for this purpose was sought and obtained from appropriate officials of the public and separate school districts of that city.

Scheduling

Actual scheduling of the experimental and control situations (presumed, by the subjects, to be critic lessons) in the schools was carried out in precisely the same way that

regular faculty consultants schedule regular observation lessons -- through the cooperating teacher who assigns lessons to the student teachers. The sole difference was that in this instance the investigator did all of the scheduling, and in person, on behalf of the Acting Faculty Consultant (experimenter) who would make the visit. This brought the cooperating teachers in on the nature and purpose of the project, a confidence which they maintained until the close of the experiment. Although a larger number of subjects was initially scheduled (approximately 90), such unavoidable complications as illness, field trips, moving days, late alterations in school schedules, and a few errors in scheduling reduced the number to 78.

REFERENCES FOR CHAPTER XI

(1) Stein, H. L. "Needs and Dimensions of Research," Phi Delta Kappan, 37 (1956) 316-317.

CHAPTER XII

PROCEDURE CONTINUED: THE EXPERIMENTAL VISIT

General Routine

The purpose of the experiment was to compare the teaching of persons submitted to stress-inducing supervision with the teaching of these same persons under more normal conditions. To be consistent with theoretical design, the experimental design had to present the subject with organizational expectations for his performance that indicated failure on his part. This had to be done in such a way that it would be need-frustrating to a subject with a strong need for achievement, ego-threatening to a subject with a strong need to maintain and enhance his self-expectations of achievement in related tasks, and generally disturbing to one who is readily disturbed. The ultimate object was to compare the experimental effect on these subjects with the effect on subjects for whom the personality descriptions are less applicable.

Immediate effects were sought. Whereas the long-range effects of prolonged stress are of considerable interest both to the theoretician and to the practicing administrator, the experiment could not humanely be so ambitious. Indeed the question of whether the experiment could be humane at all will be taken up in a later section of this chapter. To observe immediate effects it was necessary to juxtapose the control situation with the experimental one, otherwise inter-

vening activity would introduce contaminating variables. For each subject, the total experiment was, therefore, limited to two normal-length classroom periods, one following the other. Period one was the control situation, period two was the experimental or stress situation. Again, to reduce the effect of non-experimental variables, cooperating teachers were asked to assign similar subject-matter content and to impress upon the student-teachers the fact that both lessons were observation lessons, i.e. that equal preparation was necessary. A check with students' lesson plans and in later conversations indicated that this advice had been heeded. It was not difficult -- preparation for observation lessons is customarily thorough.

Student-teachers had no prior knowledge of the experimental nature of this project, nor even that it was a special project of this sort. They were not aware that ratings of these lessons did not count toward their final mark as is customary -- nor that there were no absolute ratings in the usual sense. Had it been otherwise, no superior-subordinate relationship could have been presumed to exist. The only prior knowledge afforded student teachers was a notice read out to them in class by their regular instructor, without comment, prior to the student teaching round in which the experiment was to take place. The fact that the notice was unsigned, i.e. that it did not come out over the signature of the Dean or Supervisor of Student Teaching, was never questioned in class. The notice, evidently, was accepted as genuine.

Although the notice was addressed to all second year students, it was read out to only those classes in which experimental subjects were present. It is not likely that this inconsistency would cause confusion nor raise suspisions since the notices were obviously designed to be read out to a group just prior to its round of student teaching which occurred at different times. A copy of the notice was provided in advance to each experimenter and to each cooperating teacher. The complete note, because of its relevance to experimental design, is reproduced on the following page, (p. 139).

The experimenter under the guise of an Acting Faculty Consultant bearing the same relationship to the student-teacher, so far as the latter knew, as a regular consultant, observed both lessons. His function in the experiment was to observe lesson one, induce stress, observe lesson two, compare the teaching of the second with the first, and extinguish stress. The cooperating teachers were not required to be absent during these lessons but, in general, they were.

Active stress induction intervened between lessons one and two. This was done by means of a standardized statement adapted to the present setting and theoretical requirements, from McClelland's categories for verbally-inducing failure-stress (1) reported earlier. The statement was beforehand memorized by the experimenter. It was designed to convey, as briefly as possible, the impression that the student's teaching was evaluated as being inconsistent with the expectations

To All Second-Year B.Ed. Students:

A slightly different plan for supervising student teaching will be tried out during this next round.

In addition to the regular lessons you teach for your regular faculty consultant, which will not be changed in any way, you will likely be asked to do two more lessons for an acting consultant. Your acting consultant will be a person other than your regular consultant and will go out from the Faculty of Education for the purpose of observing these lessons only.

Your co-operating teacher in the school will inform you of the name of your acting consultant and the time of his visit. For convenience this is being schedule entirely through your co-operating teacher.

These will be two consecutive lessons, both taught during one visit of your acting consultant. During this one visit you will teach one lesson followed directly by another lesson. Please let your acting consultant know when you have completed the first of these two lessons in case he wishes to discuss it before you proceed with the second lesson.

Note: To give the student a fair change, fellow student teachers are asked not to sit in on these lessons. That is, would you please be absent from the classroom when your fellow student teacher is teaching there. It is important that you do a good job of these lessons.

Notice Read to Subjects

held by the institution for the teaching activity of students in the program in which he was registered, that such failure was personal rather than situational or casuistic, that it was general, and that a fuller discussion of the student's failures would follow the second lesson. The statement was, of course, totally false. It was intended to induce stress and establish the acting consultant as a stress agent so far as this student teacher was concerned. The consultant's continued evaluational presence in the classroom after the statement was made, together with the promise (- threat?) of a fuller discussion later, may be presumed to extend the stress situation for the duration of lesson two. That there was probably some superior-subordinate stress present during the first or control lesson as well, because of the evaluational nature of all consultants' visits to student-teachers, cannot be dismissed. However, this is a study of the effects on teacher behavior of a marked increment in the objective drive stimulus.

At the close of lesson two, of course, there was no such discussion as had been promised. Rather, there was a complete disclosure of the nature and purpose of the experiment, a sincere apology and expression of gratitude for having unwittingly helped in an important research project, some value of this project to the student both directly and indirectly, assurance that nothing from this visit went on to university records and that no ratings were given. This was given conversationally, in a friendly, supporting manner, with free

student discussion. The students' further cooperation was sought in three ways: (1) they were asked for a verbal expression of how they felt during the stress period, (2) they were asked to maintain security until the close of the project at a given date, and (3) they were asked whether or not they had been informed of this experiment prior to the visit.

After the entire project was complete, the investigator made a personal announcement to each class in which subjects were registered. The purpose of this announcement was to: (1) let them know that security was lifted, (2) corroborate the stress-extinguishing statements made by experimenters, (3) describe the nature and purpose of the experiment and its value to them, and (4) provide an opportunity for questions and discussion.

Instructions to Experimenters

The foregoing general description of experimental design raises questions of standardizing the procedure followed by fifteen experimenters, particularly with respect to stress stimuli and criterion measures (D-scores). A lack of uniformity in either of these would invalidate the study. Two devices were planned the better to acquaint the experimenters with procedure and the necessity for standardization. One was the "Experimenter's Kit" containing all necessary materials and instructions for each visit; the other was a series of training sessions conducted by the present investigator.

Because the contents of the kit reveal, in greater

detail than a general statement, the specific experimental procedure at each turn, they are presented intact as follows.

Experimenter's Kit. The kit contained eight items:

1. A copy of the notice read to student teachers, presented above.

2. The individual schedule for the experimenter showing the time and place of each visit, the subjects to be visited, and orientation information. One day's entry on a specimen schedule is reproduced as a sample:

SCHEDULE: (NAME OF EXPERIMENTER)

Friday,
Feb. 24
9:15 a.m.

Holyrood School (pronounce as holly; rhymes w. mood)
7920-94 Ave.: Take Whyte Ave. to 79 St., turn N.
to school) Note: B.T.K. was there last round.

Cooperating Teacher: Mrs. S. M. Colton, Gr. I, Rm. 11.
(Walk left as you enter the bldg., to end of hall,
then right a few doors)

Faculty Consultant: Miss J. A. Y........

Student Teachers: 9:15 Miss D. M. G........
10:45 Miss M. C. F........

3. One copy, for each subject visited, of the White Rating Sheet, together with a specimen of what a completed sheet might look like. A specimen copy of a completed sheet appears below.

4. Two copies, for each subject visited, of the "Confidential Report on Student Teaching, Form 60-2", known simply as the "Pink Card". A copy of this appears in the Appendix.

5. A page of general routine, the wording of which is as follows:

"GENERAL ROUTINE: Your present kit contains all that you need.
"Go to the room, knock, take a chair at the rear of the room. See that no other student-teachers (S.T. for short) are present; cooperating teacher may or may not be present as she wishes. Do nothing during lesson one to arrouse apprehension; you should be pleasant and supportive in manner.
"During lesson 1 make notes on a pink card to later help you compare lesson 2 with this lesson.
"When lesson 1 is completed (perhaps while pupils are finishing seatwork) draw S.T. aside to verbally induce failure stress (see directions). Do not do this in front of the children; better to step into the hall. If S.T. asks questions or initiates discussion, point out that you will have a full discussion at the end of lesson 2.
"(If, as a result of your comments, S.T. indicates that she cannot proceed, do not ask her to do so. Instead, switch to the closing-out (see directions). Make a note of this on the white rating sheet and make no ratings.)
"During lesson 2, compare performance with lesson 1. Use a pink card to make notes on lesson 2 and compare notes. Indicate your comparisons in the lesson 2 rows of the white rating sheet (see directions).
"At the close of lesson 2, close out the experiment (see directions).
"Please jot down any verbal reactions of S.T. to your stress comments on the reverse on the white rating sheet.

6. A page of directions to induce failure-stress:

"DIRECTIONS FOR INDUCING FAILURE-STRESS:

"Be neither angry nor supportive, but rather emotionless as though you were stating facts which she should know. You are, in effect, making known the organization's expectations for the role of this individual -- but in a negative way.

"Do not depart from the standardized statement that follows. This is important. If you can't remember it, jot it down on the back of a pink card but don't read off this typewritten page. This statement is adapted from McClelland's categories for verbally-inducing failure stress used in several experiments. On the other hand, do not expect that this statement will produce any change in behavior necessarily, nor that the change, if any, will be negative; it may go either way or not at all. The statement follows:

> "Miss ------, I'd like to make a comment about the lesson you have just taught. You are in the degree program, aren't you?
>
> "In the Bachelor of Education program we expect a high standard of student teaching. Your lesson fails to measure up to these standards. That is, you would have to receive a failure rating for this lesson.
>
> "These lessons are rather important, you know. For your sake, they help to determine your success or failure at university and your suitability for teaching. For the sake of the pupils, they are deprived of regular instruction from their own teacher when you take over. That's why I want you to be aware of this now, before you go on with the next lesson.
>
> "You see, there are quite a number of things that you should watch out for. Your last lesson was disappointing on each of several aspects; in fact, all those things that really go to make up good teaching.
>
> "But you go ahead with the next lesson now and we'll go over these points afterward."

"(This takes about two minutes.)"

7. A page of directions for rating performance change and arriving at D-scores, which follows:

> "DIRECTIONS FOR RATING:
>
> "Your purpose is to show the degree of change in teaching behavior from lesson 1 to lesson 2. You will do this by deciding whether lesson 2 was the same, slightly better, slightly poorer, etc., and making a check mark in the lesson 2 row of the white rating sheet to indicate this.
>
> A check mark is already placed in the zero box of the lesson rows of the white rating sheet. This zero simply means a base point indicating degree of lesson one effectiveness.
>
> "Please place a check mark in the lesson 2 row to show the degree to which the effectiveness of lesson 2 departs from the effectiveness you observed during lesson 1; that is, from the zero point. Select the appropriate box like this:

+3 means marked improvement over lesson one
+2 means just some improvement over lesson one
+1 means very slight improvement over lesson one
0 means no change from lesson one
-1 means very slight deterioration from lesson one
-2 means just some deterioration from lesson one
-3 means marked deterioration from lesson one.

"Assume that the intervals between categories are equal, for example that +2 equals just as much more improvement than +1, as +1 does than 0.

"To assist in making these comparisons you may want to make some notes while lessons 1 and 2 are in progress. Your pink cards, containing the same headings and definitions as the white rating sheet, are for that purpose.

"Assume that the component attributes (Personal Qualities, Presentation of Lesson, etc.) are independent of each other when rating. Otherwise you may slip into the halo effect -- a string of similar ratings for no other reason than failure to make independent judgments.

"Do not assume that, as a result of your verbal comments or anything else that transpires during your visit, your student teacher should be expected to react in any one way or another. Please be as objective as possible when comparing the teaching of the two lessons.

8. Two pages of directions for closing out the experiment, which follow:

"DIRECTIONS FOR CLOSING OUT THE EXPERIMENT:

"You have a little more latitude here in your manner of phrasing the content. My wording below is suggestive only. But be certain to cover all these points and only these points, preferably but not necessarily in this order:

1. It was all an experiment and only so; last comments were false.
2. State purpose of the experiment.
3. Indicate value to them of this project: supervision, teaching.
4. Assure that this does not affect their record here at all.
5. Assure confidentiality, that you will not discuss it with others.

6. Give supportive assessment of the lessons observed.
7. Seek cooperation in keeping it quiet until February 24th.
8. Ask if they had been informed of the experiment beforehand.
9. Ask for verbal reactions, if they care to provide them.

"Suggested wording:

"Miss-----, I have something entirely different to say now. Your performance was absolutely wonderful. You see, these lessons were set up as part of an experimental project in supervision being carried out in the Faculty under Mr. Brown's direction. Everything I said to you was false. Completely false -- in fact, it was all written out beforehand.

"We did it on purpose so that we could learn something about the different kinds of effects that some supervision has on classroom teaching. We know that teachers sometimes have to face severe methods of supervision and other uncomfortable situations from people who may think it their job. I hope you don's meet any of this sort; you might, but different people react differently to the same comments. That's why for research purposes we had to make them all the same: many of your fellow students will be told the same thing, right to the letter.

"From what we learn about what happens to teaching under stressful conditions like this, we can try to improve supervisory methods of principals, superintendents and so forth. So, you, see, this project can have some very real value to you as a teacher. Furthermore, think of what it means for teaching: if you found it harder to teach under such conditions, how much harder must it be for a pupil to learn when his teacher creates tensions in the class.

"Now, this is important: you won't get any adverse rating on these lessons. We were just looking for a comparison between the two lessons so there is no actual rating on either. I certainly won't be discussing it with anyone, and the results won't be known to anyone excepting Mr. Brown who will only have it in coded form on cards for analysing the data from the whole group. None of the professors have access to any individual results.

"But even so, if you want my own opinion for what it is worth, I thought you did an outstanding job under these conditions. You should do very well as a

teacher (elaborate if necessary).

"This project is being carried out with most (or, 'a very large number', depending on their class) of your fellow-students. But if they find out anything that was told to you this morning/afternoon all the value of this undertaking will be lost. We can't learn anything about how a teacher teaches under stress and tension if she already expects it and knows it is all a phoney.

"So we are asking for your cooperation in this project in not mentioning anything at all about this morning's/afternoon's visit to any of the others until after it is all over, February 24th. At that time there will be a big announcement telling everyone what I've told you now. Up to now, all the student-teachers visited have agreed to co-operate in this manner, will you?

"Tell me, had you heard anything about it before?

"How did you feel about my earlier statement, did it affect you?

Experimenter Training. A series of three sessions was conducted by the present investigator with the experimenters. The first two were brief with the purpose of general orientation and discussion to develop an understanding of the basic notions of this thesis, the problem it sought to answer, and its relation to their own work both as school administrators and in their graduate program. Procedural matters were discussed in general terms.

The final and more extensive session was held a few days prior to actual experimentation. At this session the kits were distributed and studied in detail. A major purpose of this session was to be certain that standardized procedure with respect to manner and attitude, stress-induction, and assessment of teaching was followed. By this point the

experimenters had gained sufficient understanding of the project that any pre-dispositions that may earlier have been entertained appeared to have been set aside in favour of a spirit of open inquiry. An example: from much general reading and general goodwill an experimenter may be forgiven for entertaining the notion that this experiment would, in most instances, result in performance decrements. The view became accepted that this was still an open question and that any such prejudgments would, in fact, prejudice the assignment of D-scores.

The experimenters, although by now familiar with the general aims of the study, were in no way informed of the hypotheses of chapter eight. They were not provided with any clues as to how the theory would predict change, nor as to the predicted distribution of D-scores.

Security

In a project of this nature which aroused a great deal of interest if not, indeed, apprehension in the large numbers of persons involved in its administration, and where there is ample opportunity for word-of-mouth information (or, the 'grapevine') reaching the wrong persons in the wrong way, it would be imperative that strict security be maintained. This is true both for the success of the study as well as for the good relations the Faculty of Education maintains with its students, it cooperating teachers, the school systems and the community at large. If one of the subjects learned of the nature of the acting consultant's visit beforehand, this

experimental visit would be of no value. If the information entered the grapevine so that several subjects learned of it, the project would have to cease. Guarding against this liklihood was difficult because not all experimental visits were conducted on the same day. Those for one round were spread over four days of one week, for another round it was a different week.

A second possibility was that a regular faculty consultant or cooperating teacher may be unsympathetic to the project and discuss it critically with a student who had been visited, thus shaking this student's confidence in the project and perhaps tending him to break security with his fellows. Any of these eventualities would be a source of concern to university officials as well as an indication of the failure of the experimental procedure.

It is gratifying to report that no such breaks in security occurred. Whether it was due to the sincere efforts of the experimenters, to the sympathetic interest shown by others when the investigator sought their cooperation, to the prevalent research-consciousness at the Faculty of Education, to the fact that student teachers are simply good people, or to all of these, the confidence was kept. The project was carried forward virtually without incident. 'Virtually' because two potential incidents received prompt attention and the possible effects on the project were averted. Both were probably the result of errors in scheduling. In one instance, a subject who had been ill on the date of the scheduled visit

was rescheduled for a date six days later. By this time, the regular consultant, thinking the experiment had gone off, discussed the project with the unwitting subject who reported this to the experimenter only after the 'stress-induction' had been attempted. The subject was dropped from the experiment. The second instance resulted from a last-minute scheduling through a very cooperative cooperating teacher but in a school where, through oversight, the principal had not been similarly informed. Although his initial objections were quickly allayed, the experiment with that subject was also cancelled.

The conclusion that no "leaks" occurred is based on more than a lack of obvious evidence to the contrary. Fortunately there were several leads into the 'grapevine' or, stated otherwise, reliable informants on the informal organization that existed within each formal group (student-teachers, cooperating teachers, and faculty consultants) were available to the investigator. These sources reported orally and informally, directly as well as indirectly through the experimenters or other intermediaries. The reports were all negative with respect to breaks in security. Evidently the project had not become a subject of 'gossip' while still in progress. In addition, of course, there was the direct oral report provided to each experimenter at the close of a visit, and the general indication given by subjects in classes during the announcement and ensuing discussion.

Typical of the student-teachers' responses to a request for secrecy were those who felt a sense of identification with the project, smiled knowingly, and said something like, "No, you can count on me," and those who saw the gleeful side and exclaimed, "You bet I won't. After going through this I'd like to see what it does to Joe!"

The Humaneness of the Project

Raising questions of how humane was the treatment runs the risk of identifying oneself with anti-vivisectionists and the sort. The researcher's typical defences are simple; 'our knowledge must progress and such is the price of progress; or 'others do it all the time, witness the praise-punishment studies in learning;' or, more simply, 'it was the only design that would test the hypotheses.' But such defences as these, albeit true, avoid the question.

It is true that each experimenter found stress-induction difficult or even painful to himself. However, none of the subjects broke down at the point of stress-induction or ensuing teaching, even though in several instances their teaching deteriorated drastically. During the closing-out phase a sense of warmth prevailed; such was the intent and such was generally the case. Several student teachers remarked on how striking an object lesson it had been to themselves, noting how difficult a pupil must find learning under teacher-induced stress. The warmth and rapport sometimes came only after a brief initial hostility. For example, when one experimenter asked how the subject felt about the experiment she

replied: "You really want to know?", and cursed the experimenter and the experiment. But she then entered into a quite interested discussion of the methodological implications of the possible results.

Three subjects did break down -- but only after having been informed that it was only an experiment and that their teaching actually was good. With any subjects who showed visible emotion, the experimenter reported it to the investigator who arranged personal interviews for the subject with the regular faculty consultant or with himself or, in two instances, both.

REFERENCES FOR CHAPTER XII

(1) McClelland, D. C. and F. S. Apicella "A Functional Classification of Verbal Reactions to Experimentally Induced Failure," Journal of Abnormal and Social Psychology, 40 (1945) 376-390.

CHAPTER XIII

DISTRIBUTION OF PERFORMANCE CHANGE UNDER STRESS

In addition to the data provided by antecedent tests (abbreviated: MPI-N, n Ach, Ac, Ai, and ACE), the experimental procedure yielded raw data in the form of 78 White Rating Sheets. Each rating sheet contained six difference ratings, or D-scores, indicating the direction and degree of change in effectiveness of teaching behavior (1) generally or globally, (2) in the use of personal qualities, (3) in the evidence of preparation for teaching this lesson, (4) in the presentation of lessons, (5) in the manner and use of personal contact with pupils, and (6) in classroom management: routine and discipline. These six are abbreviated thus: G, PQ, PT, PL, CP and CM.

Confidence Coefficient = .95

It was decided at the outset that all statistical inferences should be based on a confidence coefficient of .95. This is often described as "working at the .05 level of significance." Doing so has implications in two directions: both "good" and "bad". A large level of significance, or alpha, exposes the investigator to the probability of committing an error of the first kind, or Type 1 Error: rejecting an hypothesis when it is true. The null hypothesis of no real difference between groups, for example, may be rejected when in fact there *is* no real difference; observed differences being accountable by chance fluctuations. But reducing this probability

immediately exposes him to the probability of committing an error of the second kind, or Type II Error: accepting the hypothesis when some alternative is, in fact, true. There is, perhaps, a true difference but not so great as would pass this more rigorous level of significance. "Most statisticians," observe Walker and Lev (2, p. 61), "do not look with favour upon choosing an extremely small level of significance because that would expose them to a large risk of error of the second kind."

The present investigation is primarily exploratory in character. In administrative relationships, the theory as presented here is new. The chief value of the study will lie in suggesting possible new ways of interpreting supervisory relationships. It is not a normative survey, nor a replication. Considering these statements leads to the suggestion that it is important for any differences existing within the subjects to be immediately indicated so that further study may be given to the variables and to the theory in general. To have 95% confidence in the findings at this stage of investigation should, therefore, be considered as quite adequate. Alpha, therefore, is set at .05 throughout.

Symbolization

To avoid confusion, the symbols that may freely be used in the text as well as in the tables are summarized here. These follow rather closely those of Walker and Lev.

The letter P will be used to denote "the probability that an obtained value will occur under the stated hypothesis." Thus, when $P > .05$, the hypothesis is accepted. When $P \leqq .05$, the hypothesis is rejected. Reference is, of course, to the operational hypothesis being tested at that moment, not to the Basic Hypothesis from which it was drawn.

The capital letter N represents the work 'number', and refers to the sample size, for any sample. A sub-sample will be denoted only by a subscript to the letter N if necessary. The letters df will refer to the number that are free to vary within a particular test, or the number of degrees of freedom. The letter f means frequency of responses within a particular D-score category.

As a rule, where necessary, Greek symbols will be reserved for reference to the parameter, Roman symbols to the sample, For example: $\overline{X}$ means sample mean, μ means the parametric mean; s and s^2 are the sample standard deviation and variance but σ and σ^2 are the parametric standard deviation and variance.

The capital letter H will refer only to operational hypotheses. When reference is to the complete statement of a basic hypothesis originally stated in chapter nine, it will be spelled out in words, for example: Hypothesis Five. The two classes of hypotheses will be related through a subscript to the letter H denoting the basic hypothesis from which it is

drawn. Thus H_5 is the null hypotheses claiming no difference between means of three groups, the high-, medium- and low-scholastic aptitude groups respectively, in their reactions to stress, whereas Hypothesis Five, in line with theory, claimed that stress would depress the mean effectiveness of the high group and enhance that of the low.

If there is need for more than one operational hypothesis to test a Basic Hypothesis, dual subscripts will be used, the first to denote the basic hypothesis and the second to distinguish between derivative operational hypotheses. Thus Hypothesis Seven may require testing by means of $H_{7.1}$, $H_{7.2}$, and so forth.

Any additional symbols that may be required will be explained prior to their use.

To avoid continual reference back to Chapter Nine, the Basic Hypotheses will be repeated where necessary.

Testing Hypothesis One:

> That teachers placed under conditions of stress-inducing supervision will, relative to control conditions, vary both in degree and direction of change in teaching effectiveness; some will deteriorate to different degrees, some will improve to different degrees, and some will remain the same.

The simplest and perhaps only necessary test of this hypothesis is a simple non-statistical observation of the distribution of D-Scores.[1] This is done with H_1 which states

---

[1] Except where otherwise stated, D-Scores will refer to the criterion score for general teaching effectiveness (g); not to components.

that D-Scores do not vary; there is no difference between any two obtained D-Scores. If a simple observation of D-Scores shows otherwise, that some scores are positive, some negative and some zero, *and if* such observation shows some positive scores to be larger than others and some negative scores to be larger than others, then H_1 is rejected in favour of Hypothesis One. Such a distribution together with statistics computed from it, is presented in Table 1.

Table 1

DISTRIBUTION OF D-SCORES[1] SHOWING FREQUENCIES AND STATISTICS CALCULATED FROM THE DISTRIBUTION, N = 78

X	f	fX	fX^2	Statistic		Value
+ 3	0	0	0			
+ 2	4	8	16	$\bar{X}$	=	- 0.38
+ 1	19	19	19	s^2	=	1.72
0	20	0	0	s	=	1.31
- 1	17	-17	17	$s_{\bar{X}}$	=	.1133
- 2	14	-28	56			
- 3	4	-12	36			
	78	-30	144			

---

[1] see footnote, page 158.

Observation of Table 1 indicates that, excepting for D-score category of plus-three, there are entries opposite each score. The effect of the experiment, therefore, was differential and H_1 may be rejected in favour of Hypothesis One.

But there is no probability level to cite with such a deduction. Another way of testing H_1 would be to compare the obtained distribution with a specified theoretical distribution and to test the comparison with a goodness-of-fit test. Such is actually the purpose of Hypothesis Two. Support for Hypothesis Two supports Hypothesis One.

First, however, the range of obtained D-scores deserves attention. Although experimenters were provided with a seven-point continuum on which to indicate change in teaching behavior, it is evident from Table 1 that only six scoring categories were used. These ranged from -3 to +2. Retaining the +3 category in the analyses that follow would add little but confusion to the tables and would not affect statistical procedures. In the following sections, therefore, the D-scores will be assumed to be a dependent variable that -- within the present investigation -- ranged on a six-point continuum extending from -3 to +2.

Testing Hypothesis Two:

That the distribution of differences will approach normality.

The operational hypothesis H_2 sets up this situation: that there is no statistically significant difference between

the obtained distribution of D-scores and a normal distribution of D-scores having μ = - 0.38 and σ = 1.31. A test of this hypothesis is the Chi-square goodness-of-fit test, data for which is shown in Table 2.

For df = 3, the probability of obtaining a chi-square value of 4.09 because of chance fluctuations is .25 (the 75th percentile of $\chi^2 = 4.1$). That is, there is one chance in four that the departure of an obtained distribution from normality will be great enough to yield a chi-square of 4.09 owing solely to the operations of chance. In this study it was decided that only when the probabilities are one in twenty (.05 level) shall we conclude that the departure may be attributed to factors other than chance. We must, therefore, conclude that the evidence is not strong enough to reject H_2: that there is no significant difference between our obtained distribution and a hypothetical normal distribution having the same mean and standard deviation. Acceptance of H_2 does not, however, automatically imply that our distribution is normal, only that it is not significantly non-normal. This information supports Basic Hypothesis Two, that it approaches normality. It will be noted that this information supports Hypothesis One, as well.

Testing Hypothesis Three:

> That the general or average tendency of teachers under conditions of supervisory tension will be in the direction of decreased teaching effectiveness.

The null hypothesis H_3 states that there is no

Table 2

GOODNESS OF FIT OF A NORMAL DISTRIBUTION WITH AN OBTAINED DISTRIBUTION OF D-SCORES SHOWING REACTIONS UNDER STRESS, N = 78[1]

D-Score	O Observed	Upper Limit	Deviation from $\overline{X}$	$\frac{x}{s}$	Proportion Below	Proportion Within	E Expected	O-E	$\frac{(O-E)^2}{E}$
+ 2	4					.0749	5.8	1.8	.559
+ 1	19	1.5	1.88	1.44	.9251	.1765	13.7	5.3	2.050
0	20	0.5	0.88	.67	.7481	.2845	22.2	2.2	.218
- 1	17	-0.5	-0.12	- .09	.4641	.2664	20.8	3.8	.694
- 2	14	-1.5	-1.12	- .85	.1977	.1610	12.6	1.4	.155
- 3	4	-2.5	-2.12	-1.69	.0367	.0367	2.9	1.1	.417
Total	78	...	...	...	...	1.0000	78.0	...	4.093*

[1] after method of Ferguson (1, p. 163)

* under H_2, $P(\chi^2 \geqq 4.1) = .25$

difference between a hypothetical mean D-score of 0 and the obtained mean D-score of -0.38. This assumes, of course that "general or average tendency" is best indicated by the mean. Other measures exist for estimating μ, such as the median, mode, or average of 25th and 75th percentiles, or the midpoint of the range, etc. However, in samples from a normal population every other estimate of μ has a standard error larger than $\sigma_{\overline{X}}$ (Walker and Lev, 1953, p. 144). Since the obtained

distribution has been demonstrated to be not significantly different from a normal distribution, the reliability of $\overline{X}$ is $\frac{s}{\sqrt{N}}$ given earlier as .1133. The reliability of the difference between this mean and a hypothetical mean of zero resolves to a probability statement about the reliability of the obtained mean, or

$$P\ (+\ 3.00\ s_{\overline{X}} > \mu > -\ 3.00\ s_{\overline{X}}) = .997.$$

That is, there are only three chances in one thousand that the true mean will lie outside three standard errors on either side of the obtained mean. Adding three standard errors (3 x .1133) to the mean (- 0.38) still does not reach 0. Working at the .05 level of significance, therefore, we reject the null hypothesis H_3 and conclude with great confidence ($P < .003$) that the obtained negative mean of -0.38 significantly departs from a hypothetical zero point. Basic Hypothesis Three is thus supported; student-teachers placed under severe conditions of negative supervision that should tend to create stress, do, as a general tendency, deteriorate in their teaching effectiveness. The deterioration is statistically significant, although relatively slight when compared with the overall range of D-scores.

The three hypotheses tested in this chapter refer only to what happens when an experimental situation is created in which student-teachers are told their work is a complete failure. Why it happens, or how it happens to some, as compared with others, is the question to be taken up in the following chapters.

REFERENCES FOR CHAPTER XIII

(1) Ferguson, G.A. Statistical Analysis in Psychology and Education. New York: McGraw-Hill, 1959.

(2) Walker, H. M. and J. Lev. Statistical Inference. New York: Henry Holt, 1953.

CHAPTER XIV

SEPARATE EFFECTS OF PERSONAL FACTORS

Three Levels of Each Factor

The reasons for dividing the experimental sample into three equal groups on each of the personal factors is two fold. The first is simply that the tests, although reasonably reliable and accurate, cannot be thought of as so precise as to discriminate with great confidence between each of, say, ten or twenty levels. Correlational analysis presumes, for example, that the instrument discriminates into as many levels as there are scoring intervals. But the general concensus of opinion is that personality tests can discriminate rather well between, say, two, three or four levels. Selecting three levels was done on the basis of the second reason: the study was designed to find out what 'kinds' of people tended to deteriorate under stress, to improve under what seemed like stress, and so forth. Selecting three levels made it possible to draw inferences regarding "people who are quite neurotic" as compared with those who are "just average in their neurotic tendencies" and with those who are emotionally very stable -- much less "neurotic" than average.

Levels Fixed, Not Random

For analyzing the data it is significant to note that the process of dividing the experimental sample into three equal groups on one basis of classification -- with no subjects

left over -- means that these three treatment levels are fixed and constant. If, however, the group had been divided into ten or twenty levels and only three -- any three -- levels picked at random from this larger number, the levels would be said to be random. Another time, the investigator may pick (again, randomly) a different set of three levels. But when equal thirds are used, there are only three possible levels. Ferguson (1, p. 250) states that, "Any factor is fixed if the investigator, upon repeating the experiment, would use the *same* levels of it."

The fact that levels are fixed, not random, is important for the selection of a "choice-of-error" term in a two-way variance design, which will be described in a later section. For the present, it is significant for the purpose of interpreting findings to know that the sample has been divided into three groups on each variable. They were not, however, exactly equal in every instance. Although the intent was to construct equal groups there were two main reasons why this was not completely realized. First of all, the necessary reduction in the experimental N from 90 to 78 because of unforeseen and fortuitous causes tended to shift the composition of the sample. Second, it turned out that the remainder did not divide evenly: two or more scores would accumulate at the cut-off point. It would be possible to equalize the subclass frequencies by *randomly* casting out an appropriate number of scores. This is wasteful and, fortunately, not necessary. For use with data that do not divide into exactly equal subclass frequencies there exist computational procedures which are, in fact, exactly the same

procedures that are used when subclass frequencies are equal, except for the fact that a little more arithmetic is required.

Simple Variance Design

One way of determining whether or not three groups respond differentially to their respective treatments is to refer the degree of dispersion[1] found among the population of group means to the degree of dispersion found among the D-scores of subjects within each group. If both measures, each of which is an estimate of the kind of dispersion one expects to find in the total population under the null hypothesis, tend to agree, then we conclude that despite differential treatments our three groups are actually drawn from a common population with respect to treatments. The null hypothesis is supported. But if they do not agree, if their ratio is a value that could have occurred under the null hypothesis in as few as 5 instances out of 100, one is forced to conclude that the use of the null hypothesis was unjustified and that some alternate hypothesis is likely the case. Depending upon observation and analysis, it is likely that the treatments did, in fact, sort the subjects into different populations with respect to the performance of the activity in question.

---

[1] "Degree of dispersion" is the writer's convenient way of denoting the concept of variance. To be more precise, variance is directly related to standard deviation which Guilford (2, p. 85) explains is an "indicator of degree of variability, . . ." and a 'dependable' and 'accurate' "estimate of the dispersion of the population.

Assumptions

One must hasten to surround the above simple statement with several qualifications: if the value did reach the .05 level as stated, all we can conclude is that one assumption has been broken. Unless it is certain that the only assumption that could have been broken was the assumption of identical sub-population means, the above conclusion is unwarranted. The assumptions of this procedure are well known. The ratio of mean squares for treatments to mean squares for within groups may be tested with the analysis of variance procedure under these conditions:

1. independence of observations
2. normally distributed populations (including sub-populations)
3. homoscedasticity of variance
4. identical sub-population means (the null hypothesis)

(3, p. 73)

To these, Siegel (4, p. 19) adds the assumption of interval equivalence required by all parametric tests. That is, one must be able to perform operations of simple arithmetic with the data. The assumption of additive effects is also made.

Independence of observations was established in the chapters on procedure. Identity of sub-population means is the hypothesis to be tested in each analysis. That sub-populations are normal and homoscedastic will be shown with each analysis either by the application of statistical tests or -- when it is possible -- by direct inspection of the data. Although one

cannot actually test the equivalence of intervals *per se*, when a variable is normally distributed the appearance of equivalence of scoring intervals may be demonstrated. Normality of distribution of the D-score was supported by the test of Hypothesis Two (Chapter Thirteen). Hence it may be shown that sigma-distances between scoring categories are similar to those of a variable that is normally distributed on an interval scale having the same limited range. It is important to recall, furthermore, that the experimenters were instructed to assume the difference between any two successive scoring categories to be equal to the difference between any two successive categories (Chapter Twelve). Ultimately, interval equivalence cannot be tested; one assumes either that the points on the rating continuum were equally spaced in the minds of the raters, or that they were not. The writer believes there is sufficient justification for the assumption that they were so. The choice of test is, therefore, the F-test for:

> When we have reason to believe that these conditions are met in the data under analysis, then we should certainly choose a parametric statistical test such as *t* or *F* for analyzing those data. Such a choice is optimum because the parametric test will be the most powerful for rejecting . . . (the null hypothesis) when it should be rejected. (4, p. 20)

Mathematical Model and Hypothesis

The mathematical population described by the above set of conditions is the one represented by the following model provided by Walker and Lev (5, p. 201):

Population	Mean	Variance
A	μ_A	σ^2
B	μ_B	σ^2
C	μ_C	σ^2

(Simple Variance Model)

which permits the hypothesis:

$$\mu_A = \mu_B = \mu_C = \mu$$

to be tested with the F-ratio. The null hypothesis implies that the means of each of the three samples are but random observations from a normal population of sample means for which the best expectation for an obtained mean is μ, with $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{N}$.

Now the unknown σ^2 can be estimated from the variation between the three obtained sample means, but it can also be estimated from the variation among subjects within each sample or sub-population. The ratio of these two variations provides the statistic F by means of which the above hypothesis can be tested.

Testing Hypothesis Four:

> That increasing the external stress stimulus through ego-threatening or failure stress will depress the effectiveness of high-neuroticism subjects and will enhance the effectiveness of low-neuroticism subjects.

The null hypothesis H_4 dictates that there is no difference between mean D-scores of three groups classified

as low-neuroticism, medium-neuroticism, and high-neuroticism respectively. The distribution of their D-scores is presented in Table 3, below, together with certain statistics drawn from the data. The statistics N, ΣX, and ΣX^2 are required for the analysis. The statistic $\overline{X}$ is not, but is provided for illustrative purposes, particularly since N's are unequal. The statistic s^2 is required for the homoscedasticity test.

Application of the Hartley test[1] of the hypothesis that

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$$

indicates that the ratio of the greatest variance to the least, or F_{max}, is 1.42. Referring this to the probability statement (from tables of F_{max}) that $P(F_{max} > 2.40) = .05$, yields the conclusion that under the null hypothesis of the last page, $P \gg .05$, the significance level for this investigation. The null hypothesis is supported and the assumption of homoscedasticity is accepted. Inspection of the distributions indicates that the assumption of normality is also warranted, without the necessity of testing each distribution with a goodness of fit test.

Inspectionally, the mean difference between the low neuroticism group and medium neuroticism group is .459; between the medium and high is .818; and between low and high is 1.277. The null hypothesis stated at the opening of the section said that there was no real difference between these,

---

[1] taken from Walker and Lev (5, pp. 191-193, and tables of F_{max}).

Table 3

DISTRIBUTIONS OF D-SCORES OF THREE SUBPOPULATIONS CLASSIFIED BY LEVEL OF NEUROTICISM, (MPI-N)

N = 78

Level:	Low Neuroticism	Medium Neuroticism	High Neuroticism
D-Score	f	f	f
+ 2	3	1	0
+ 1	8	9	2
0	7	7	6
- 1	5	4	8
- 2	2	4	8
- 3	0	2	2
N	25	27	26
$\overline{X}$	+ 0.200	- 0.259	- 1.077
s^2	1.33	1.89	1.36
ΣX	+ 5	- 7	-28
ΣX^2	33	51	60

in statistical terms. To test this by analyzing the variance the following data are provided in Table 4.

Table 4

ANALYSIS OF VARIANCE OF D-SCORES FOR THREE SUBPOPULATIONS CLASSIFIED BY LEVEL OF NEUROTICISM (MPI-N)[1]

Source of Variation	Sum of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Total	132.46	77				
Between Means	21.42	2	10.71	7.24	3.12	4.90
Within Groups	111.04	75	1.48			

Table 4 provides the information that the 99th percentile of the F-distribution for appropriate degrees of freedom is a value of 4.90. In actual fact, an obtained F-ratio of 7.24 as shown in the table, falls at the 99.9th percentile approximately ($F_{.999} = 7.46$). Therefore working at the .05 level, we reject the null hypothesis and conclude with great confidence ($P = .001$) that the means on change-in-effectiveness scores of three groups of student-teachers differ significantly, those with a higher degree of neuroticism deteriorating more than those with a lower degree of neuroticism, and those in the mid-range falling between the two. In so far as this experiment was concerned, the three groups cannot be considered to belong to a common population of neuroticism tendencies in their manner of responding to treatment. It is interesting to note that the middle group did deteriorate slightly, in line with Basic Hypothesis Three, tested in Chapter Thirteen. Basic Hypothesis Four, which opened this section, is now supported.

---

[1] after the computational procedure when subgroups are not of uniform N (5 p. 212).

Whether or not the characteristic of neuroticism may be termed a "need-disposition" may, by this time, be "flogging a dead horse." Nevertheless it is clear that the reactions of student-teachers under conditions where the normative and personal dimensions conflict do vary from a few who improve to a number who deteriorate, and that the personal dimension includes the emotional arousibility of the subject which has a very powerful effect upon subsequent performance.

Testing Hypothesis Five:

> That increasing the external stress stimulus through ego-threatening or failure stress will depress the effectiveness of subjects evidencing experience of high scholastic aptitude and will enhance the effectiveness of subjects evidencing experience of low scholastic aptitude.

This hypothesis is tested in exactly the same way as the previous one and will thus require considerably less elaboration regarding procedure. The null hypothesis H_5 dictates that there is no difference between mean D-scores of three groups classified as low scholastic aptitude, medium scholastic aptitude and high scholastic aptitude respectively. That is, that $\mu_1 = \mu_2 = \mu_3 = \mu$.

The distribution of their D-scores is presented in Table 5 on page 175.

The Hartley test for homoscedasticity applied to these variances yields $F_{max} = 1.47$ which, using the same probability statement as previously (p.171), permits us to accept the null hypothesis ($P > .05$) and conclude that the

Table 5

DISTRIBUTIONS OF D-SCORES OF THREE SUB-POPULATIONS CLASSIFIED BY LEVEL OF SCHOLASTIC APTITUDE (A.C.E.) $N = 78$

Level:	Low A.C.E. (1)	Medium A.C.E. (2)	High A.C.E. (3)
D-Score	f	f	f
+ 2	4	0	0
+ 1	6	8	5
0	9	4	7
- 1	3	9	5
- 2	4	5	5
- 3	0	0	4
N	26	26	26
$\overline{X}$	+ 0.115	- 0.423	- 0.846
s^2	1.63	1.29	1.89
ΣX	+ 3	-11	-22
ΣX^2	41	37	66

three variances of Table 5 are homogeneous.

In the event that some doubt as to the assumption of normality of each of these three distributions should arise, it is well to focus on the one appearing to be the least nearly normal, i.e. Medium A.C.E., distribution number 2. Testing the

hypothesis that Distribution 2 is normal, states the null hypothesis that there is no difference between distribution 2 and a normal distribution. An appropriate test of this hypothesis is the Kolmogorov-Smirnov one-sample test,[1] which requires measurement in at least an ordinal scale and requires measurement on a continuous variable (both of which assumptions are met in this instance). Data necessary for this test are supplied in Table 6, below, in which the columns represent D-score categories; rows represent, respectively, the cumulative proportions expected under a normal curve, the actual frequencies from Distribution 2, cumulative proportions of N_2 from Distribution 2, and the differences between two sets of proportions. The largest difference is the statistic D required for the test; for illustrative purposes, it is placed in parentheses.

Table 6

GOODNESS OF FIT OF NORMAL DISTRIBUTION WITH AN OBTAINED DISTRIBUTION OF D-SCORES FROM 26 SUBJECTS CLASSIFIED AS MEDIUM-LEVEL A.C.E.

D-Score:	- 3	- 2	- 1	0	+ 1	+ 2
Expected:	.0228	.1587	.5000	.8413	.9772	1.0000
f:	0	5	9	4	8	0
Observed:	.0000	.1923	.5384	.6923	1.0000	1.0000
Difference:	.0228	.0336	.0384	.1490	.0228	.0000

---

[1] after the method presented by Siegel (4, p. 47) and tables of D.

Table 6 indicates that D = .1490. Statistical tables show that for N = 26, the probability statement under the null hypothesis is as follows: $P\ (D > .21) = .20$. P varies inversely with N and inversely with D. Hence, for an obtained D of .1490, $P \gg .05$. Thus we accept the null hypothesis and conclude that Distribution 2 of Table 5 does not depart significantly from a normal distribution and, by inference, neither do Distributions 1 or 3. It is, therefore, admissible to proceed with the analysis of variance of the D-Scores obtained by the present classification of subjects, which is presented in Table 7, below, to test the significance of the obtained mean differences:

$$\bar{X}_2 - \bar{X}_1 = .538,\ \bar{X}_3 - \bar{X}_2 = .423,\ \bar{X}_3 - \bar{X}_1 = .961.$$

Table 7

ANALYSIS OF VARIANCE OF D-SCORES FOR THREE SUBPOPULATIONS CLASSIFIED BY LEVEL OF SCHOLASTIC APTITUDE (A.C.E.)

Source of Variation	Sum of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Total	132.46	77				
Between Means	12.07	2	6.04	3.78	3.12	4.90
Within Groups	120.39	75	1.60			

Table 4 shows that the obtained F-ratio of 3.78 is greater than the F-ratio falling at the 95th percentile of the distribution of F which is required when working at the .05 level. In actual fact, 3.78 falls at the 96.5th percentile.

Therefore we reject the null hypothesis and conclude with confidence (P = .035) that the means on change-in-effectiveness scores of three groups of student teachers differ significantly, those with higher scholastic aptitude scores deteriorating more than those with low scholastic aptitude scores, and those in the mid-range falling between the two. Again, those in the mid-range did deteriorate slightly.

Evidently the subject's level of scholastic aptitude (or, as the test manual also defines it, general intelligence) does enter into the total drive arousibility level which, if increased markedly through experimental stimulation, tends to have an effect upon performance on the complex activity of teaching. This effect is in the direction of greater or lesser effectiveness, though usually lesser, depending upon whether the subject is quite bright or less so. To explain this phenomenon one may adduce the theory, from Chapter Eight, that the level of measured scholastic aptitude reflects an individual's self-expectations or "set" toward the task. The degree to which the subject would find the 'stress stimulus' consistent or inconsistent with his self-image would be the degree to which he would perceive the situation as ego-threatening.

Testing Hypothesis Six:

> That increasing the external stress stimulus through ego-threatening or failure stress will depress the effectiveness of subjects possessing a strong need for achievement and will enhance the effectiveness of subjects possessing a weak need for achievement.

Because there are three extant measures of this variable or its aspects, all three were included under the operational definitions of Chapter Ten. Here, they will be tested separately in three one-way analysis of variance tests.

Need for Achievement

The null hypothesis $H_{6.1}$ dictates that there is no mean difference between mean D-scores of three groups classified on the basis of the n.Ach scale of the EPPS as having a strong need for achievement, respectively. That is $\mu_1 = \mu_2 = \mu_3 = \mu$.

The distribution of their D-scores is presented in Table 8 on page 165. Considering that the three obtained variances are so nearly identical (1.79, 1.72, 1.81) and are considerably more homogeneous than the previous two sets tested, the Hartly test for homoscedasticity is unnecessary; the assumption is justified on the basis of observation and comparison. The same is true with respect to normality of distributions: all appear more nearly normal than the distribution previously tested with the Kolmogorov-Smirnof test.

It is, therefore, permissible to test the significance of the differences $\overline{X}_2 - \overline{X}_1 = .088$, $\overline{X}_3 - \overline{X}_2 = .132$, and $\overline{X}_3 - \overline{X}_1 = .044$, using the one-way analysis of variance in Table 9, below.

Considering the exceedingly small mean differences observed in Table 8 (and reported above), it is not surprising that the F test of the null hypothesis yields an overwhelming

Table 8

DISTRIBUTIONS OF D-SCORES OF THREE SUBPOPULATIONS CLASSIFIED BY LEVEL OF NEED FOR ACHIEVEMENT (N.ACH), N = 78

Level:	Low n.Ach (1)	Medium n.Ach (2)	High n.Ach (3)
D-Score	f	f	f
+ 2	1	1	2
+ 1	6	7	6
0	5	10	5
- 1	5	5	7
- 2	4	6	4
- 3	1	2	1
N	22	31	25
$\bar{X}$	- 0.364	- 0.452	- 0.320
s^2	1.79	1.72	1.81
ΣX	- 8	-14	- 8
ΣX^2	40	58	46

support for the theory of "no-differences". To verbalize it, it would appear that three groups of student-teachers classified into three levels of need for achievement on the basis of their scores on the EPPS n Ach. scale, exhibit no difference in the effect of stress-supervision upon their classroom teaching behavior. Some improve, some deteriorate, and some remain the

Table 9

ANALYSIS OF VARIANCE OF D-SCORES FOR THREE SUBPOPULATIONS CLASSIFIED BY LEVEL OF NEED FOR ACHIEVEMENT (n Ach)

Source of Variation	Sum of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Total	132.46	77				
Between Means	0.25	2	0.13	< 1.00	3.12	4.90
Within Groups	132.21	75	1.76			

same, within each class. The total experimental sample of 78 subjects may, therefore, be considered as drawn from a common population with respect to the "treatment" of need for achievement.

Implications of $F < 1$

But before concluding thus, it would be well to examine the implications of an obtained F of less than unity. In most experiments, they are rare and lead statisticians to offer these explanations (these are taken from 5, p 207, but are typical of other sources):

1. the computations may be incorrect,
2. the data may be artificial or improperly gathered
3. the subjects were not drawn at random from the population with respect to the treatment under consideration, or
4. it is simply the operations of chance.

The first does not apply; computations were checked and re-checked by another person. They are accurate. Nor does the second apply; the data are real and were conscientiously gathered. It is possible that the fourth statement is correct, that one should expect chance fluctuations to produce very small F-ratios now and again, although rarely. Some doubt is cast upon this explanation by the succeeding two sections of this chapter. Although it is anticipating these sections, it will be shown that all analyses based on need-for-achievement variables yield F-ratios of less than unity. The compelling interpretation is, therefore, that cases were not selected at random with respect to this particular variable.

This explanation seems not at all implausible. The characteristics of the sample, stated in Chapter Eleven, include a large number of common factors among the subjects that would tend to indicate their relative homogeneity with respect to need for achievement. It is ironical that the attempt to achieve a homogeneity that would control a number of non-experimental variables probably resulted in a sample that was too homogeneous for the free operation of the particular experimental variable of need for achievement. Comparing the experimental sample with a random group from the entire population of persons of this age-group from all walks of life would lead to the inference that the former possesses a higher and more consistent desire to achieve success in school-type pursuits than would be the overall characteristic of the population. In addition to this, the plan of the experimental procedure probably resulted in an increased uniformity of need for achievement, at least situationally. The subjects were placed in observation

lessons, were told that it was important to do a good job, and -- as with observation lessons generally -- saw this as a critical situation demanding a high level of motivation. The inference is drawn by the present investigator that both selective and situational factors hindered the free operation of the need-for-achievement variable in the experimental sample so that there was insufficient opportunity for the effects of the stress stimulus to discriminate between different levels. The "levels" were simply not that different.

But if this is the explanation, what are the implications -- to theory and to design? To theory, the acceptance of the null hypothesis $H_{6.1}$ with respect to the n. Ach (and, as will be demonstrated shortly, to the other need-for-achievement variables as well) casts doubt on Basic Hypothesis Six. To design, however, the implications are even more serious. It would appear that, with respect to this particular variable, a basic assumption of the F-test has been broken: that the three treatment groups were drawn at random from an hypothetical population of three treatment groups (Independence of observation). If this is true then the advisability of using the F-test (or, for that matter, any test intended to discriminate between levels that perhaps do not exist in the experimental sample) is in doubt. To infer one more step: if the test is inappropriate, then it can neither support nor reject the Basic Hypothesis Six (or $H_{6.1}$).

It may still be true, as suggested by the theory, that the effects of stress-inducing supervision vary partly with the

subject's need for achievement but the present investigation has failed to substantiate it.

The reason for inserting the foregoing discussion prior to presenting analyses of D-scores for subjects classified according to each of "need for achievement via conformity" and "need for achievement via independence" is clear. When an investigator learns of an error in design, he must at that point report the finding and make plans for altering the design. The decision here is to present the two additional analyses for illustrative purposes, but to drop the variable of need-for-achievement from subsequent analyses.

Achievement via Conformity

The null hypothesis $H_{6.2}$ derived from Basic Hypothesis Six would dictate that there is no mean difference between subjects classified as possessing a strong need to achieve via conforming, a medium-strength need and a weak need of this nature. The distributions of D-scores assigned to the student-teachers so classified are presented in Table 10 on page 185.

Table 10 shows variances that, when Hartley's test for homogeneity of variance is applied, yield an F_{max} of 1.35 which is considerably smaller than the value of 2.40 required to reach the .05 level of significance selected for this investigation. Hence they are homogeneous. The distributions are more nearly normal than an earlier one which did not significantly depart from normality. The mean differences,

$\overline{X}_2 - \overline{X}_1 = .403$, $\overline{X}_3 - \overline{X}_2 = .269$, and $\overline{X}_3 - \overline{X}_1 = .016$, are not large.

Table 10

DISTRIBUTIONS OF D-SCORES OF THREE SUBPOPULATIONS CLASSIFIED BY LEVEL OF NEED FOR ACHIEVEMENT VIA CONFORMITY (Ac), N = 78

Level:	Low Ac (1)	Medium Ac (2)	High Ac (3)
D-Score	f	f	f
+ 2	1	1	2
+ 1	6	8	5
0	5	8	7
- 1	6	5	6
- 2	7	3	4
- 3	1	1	2
N	26	26	26
$\overline{X}$	- 0.577	- 0.154	- 0.423
s^2	1.77	1.43	1.93
ΣX	-15	- 4	-11
ΣX^2	53	38	53

The fact that these mean differences are not large is amply illustrated by the obtained $F < 1$ shown in the analysis of variance, Table 11.

Table 11

ANALYSIS OF VARIANCE OF D-SCORES OF THREE SUBPOPULATIONS CLASSIFIED BY LEVEL OF NEED: ACHIEVEMENT VIA CONFORMITY (Ac)

Source of Variation	Sum of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Total	132.46	77				
Between Means	2.38	2	1.19	< 1.00	3.12	4.90
Within Groups	130.08	75	1.73			

The interpretation of the information provided in Table 11 has already been anticipated in the discussion of the previous section and no further comment is necessary excepting to note that $H_{6.2}$ is supported.

Achievement via Independence

In a like manner, $H_{6.3}$ dictates that there is no mean difference between the D-scores assigned to student teachers for their change in effectiveness under stress, when the subjects are classified on the basis of their Ai scores into three levels: strong, medium, and weak need to achieve via exercizing or expressing independence. The distributions of these D-scores appear in Table 12. By inspection of the data it will be seen that these variances are more nearly homogeneous than those of Table 10 which were accepted as homogeneous under the Hartley test. Although distributions 1 and 3 of Table 12 do not appear to conform to the concept of the normal probability curve, both

more nearly conform than distribution 2 of Table 5 which was shown in Table 6 to depart from normality to a degree that was not significant.

Table 12

DISTRIBUTIONS OF D-SCORES OF THREE SUBPOPULATIONS CLASSIFIED BY LEVEL OF NEED: ACHIEVEMENT VIA INDEPENDENCE (Ai), N = 78

Level:	Low Ai (1)	Medium Ai (2)	High Ai (2)
D-Score	f	f	f
+ 2	2	1	1
+ 1	5	8	8
0	6	7	7
- 1	5	5	5
- 2	8	5	1
- 3	1	1	2
N	27	27	24
$\overline{X}$	- 0.556	- 0.297	- 0.292
s^2	1.33	1.49	1.60
ΣX	-15	- 8	- 7
ΣX^2	43	41	39

Not only are the mean differences in Table 12 slight ($\overline{X}_2 - \overline{X}_1 = .261$; $\overline{X}_3 - \overline{X}_2 = .005$; $\overline{X}_3 - \overline{X}_1 = .266$), but they run in a direction directly counter to theory. Because they are not significant, however, this observation warrants no

further comment. Their lack of significance is demonstrated by the analysis of variance (see Table 13).

Table 13

ANALYSIS OF VARIANCE OF D-SCORES OF THREE SUBPOPULATIONS CLASSIFIED BY LEVEL OF NEED: ACHIEVEMENT VIA INDEPENDENCE (Ai).

Source of Variation	Sum of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Total	132.46	77				
Between Means	1.20	2	0.60	<1.00	3.12	4.90
Within Groups	131.26	75	1.75			

The null hypothesis $H_{6.3}$ is supported and it is concluded that the experimental sample N = 78 may be assumed to be drawn from a common population with respect to the generalized need for achievement which has been tested by means of three scales: n Ach, Ac and Ai.

Chapter Fourteen has served to support Basic Hypotheses Four and Five, but not Six, in so far as these variables were investigated.

REFERENCES FOR CHAPTER FOURTEEN

(1) Ferguson, G. A. *Statistical Analysis in Psychology and Education*. New York: McGraw-Hill, 1959.

(2) Guilford, J. P. *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill, 1956.

(3) Lindquist, E. F. *Design and Analysis of Experiments in Psychology and Education* (Boston: Houghton-Mifflin, 1953.

(4) Siegel, S. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill Book Co., 1956.

(5) Walker, H. M. and J. Lev. *Statistical Inference*. New York: Henry Holt, 1953.

CHAPTER XV

NEUROTICISM AND SCHOLASTIC APTITUDE: INTERACTION AND MAIN EFFECTS

In Chapter Fourteen, two of the drive arousibility variables were shown individually to produce the hypothesized changes in a student teacher's behavior when these variables were combined with a marked increment in the objective drive stimulus. In each analysis, however, it must be assumed that the contributing effect, if any, of the other factor was uncontrolled. In the absence of evidence showing either that each factor contributed independently or that the effect of certain levels of one factor was greater than with certain levels of the other, the findings must be regarded as tentative.

A two-way factorial analysis permits the investigator to observe the main effect of each factor independent of the effects of the other factor, and to determine whether or not an interaction effect is present. An interaction effect, in the present investigation, would occur where, for example, criterion means for low-neuroticism subjects were equal for each level of scholastic aptitude. In this instance, neuroticism interacts with scholastic aptitude by nullifying the latter's effect at low levels of the former. Interaction may occur in any or all cells; because of chance fluctuations, it would be unusual for an observed interaction to equal zero. Were interaction zero, the means for the low neuroticism level, to continue

the example, would be exactly proportional to the overall means for scholastic aptitude. Zero interaction in a two-way layout implies that cell means are determined entirely by the marginal means. The two variables function independently; their effects are additive. This is the situation that was predicted in Basic Hypothesis Seven.

Mathematical Model and Hypotheses

In a two-way design with replication, the total variance may be partitioned into four additive components: the variance for rows, for columns, for interaction, and for within cells. Each is assumed to be an efficient estimate of the population variance.

The particular two-way factorial design with replication used in the present investigation is based on a mathematical model, the population values of which are estimated in the symbolic analysis of Table 14, where r, c, and n denote number of rows, columns and individuals per cell respectively, and where subscripts represent the *i*th row, *j*th column, and αth individual in a cell.

In this model, the mean of each cell is assumed to be the constant $\mu_{ij.}$ Estimated population values are determined only on this assumption. Each of the null hypotheses

H_r : $\mu_{1.} = \mu_{2.} = \mu_{3.} = \mu$ (that row means are equal),

H_c : $\mu_{.1} = \mu_{.2} = \mu_{.3} = \mu$ (that column means are equal), and

H : $\mu_{ij} - \mu_{i.} - \mu_{j.} + \mu = 0$ (that interaction is zero)

Table 14

POPULATION VALUES ESTIMATED BY MEAN SQUARES IN AN EXPERIMENT WITH $\underline{r}$ ROWS, $\underline{c}$ COLUMNS AND $\underline{n}$ INDIVIDUALS IN EACH CELL[1]

Source of Variation	Mean Square	Population Value Estimated
Rows	$\frac{cn}{r-1}\Sigma(\bar{X}_{i.}-\bar{X})^2$	$\sigma^2 + \frac{cn}{r-1}\Sigma(\mu_{i.}-\mu)^2$
Columns	$\frac{rn}{c-1}\Sigma(\bar{X}_{.j}-\bar{X})^2$	$\sigma^2 + \frac{rn}{c-1}\Sigma(\mu_{.j}-\mu)^2$
Interaction	$\frac{n\Sigma\Sigma(\bar{X}_{ij}-\bar{X}_{i.}-\bar{X}_{.j}+\bar{X})^2}{(r-1)(c-1)}$	$\sigma^2 + \frac{n\Sigma\Sigma(\mu_{ij}-\mu_{i.}-\mu_{.j}+\mu)^2}{(r-1)(c-1)}$
Individuals Within Subclasses	$\frac{\Sigma\Sigma\Sigma(X_{ij}-\bar{X}_{ij})^2}{rc(n-1)}$	σ^2

---

[1] from Walker and Lev (8, p. 361)

modifies the population values so that each reverts simply to σ^2 because each of the summed deviation values, e.g. $\Sigma(\mu_{1.}-\mu)^2$, becomes zero. Thus, each F-ratio satisfies the condition that the numerator and the denominator are estimates of the same variance.

Although the above symbolic analysis forms the basis of both a fixed-constants model and a random-replications model, the choice-of-error term for testing the null hypothesis differs. It was earlier demonstrated that, in the present investigation, the variables are classified as fixed, rather than random levels drawn from among many. In such a model the appropriate choice-

of-error term to be placed in the denominator of the F-ratio is that of the variance of individuals within subclasses. (The random model employs interaction variance.) (3, p. 251)

The usual assumptions of homoscedasticity and of normality of distributions apply to each of the individual cells in a two-way layout, rather than to total distribution on each level of a variable.

Obtained Two-Way Layout and Assumptions

The data used in the two-way analysis of variance are identical to those entered into the individual one-way tables of Chapter Fourteen. This time they have been classified on both variables simultaneously, rows representing levels of neuroticism (MPI-N) and columns representing levels of scholastic aptitude (ACE). These data appear in Table 15.

Table 15 presents nine different subpopulations into which the experimental sample (N = 78) has been divided. Each subpopulation is described by its respective row and column designation, Cell_{11} being composed of subjects whose neuroticism scores are high but whose ACE scores are low, Cell_{12} being those with High MPI-N and medium ACE scores, and so on to Cell_{33} being those with Low MPI-N and High ACE scores.

To test the assumption of homogeneity of variance, two statistics not included in Table 15 are required for each cell: its sum of squares and its variance. For illustration, these are entered into Table 16, which retains the same

Table 15

D-SCORES OF 78 STUDENT TEACHERS SUBMITTED TO STRESS-INDUCING SUPERVISION, CLASSIFIED BY LEVEL OF NEUROTICISM (MPI-N) AND BY LEVEL OF SCHOLASTIC APTITUDE (ACE)

Level of Neuroticism (MPI-N)	X	Level of Scholastic Aptitude (A.C.E.) Low	Medium	High	Totals
High Neuroticism	+2	0	0	0	
	+1	1	0	1	
	0	4	1	1	
	-1	2	5	1	
	-2	3	2	3	
	-3	0	0	2	
	N	10	8	8	26
	ΣX	- 7	- 9	-12	-28
	$\bar{X}$	- 0.70	- 1.13	- 1.50	- 1.08
Medium Neuroticism	+2	1	0	0	
	+1	2	4	3	
	0	4	1	2	
	-1	0	3	1	
	-2	1	2	1	
	-3	0	0	2	
	N	8	10	9	27
	ΣX	+ 2	- 3	- 6	- 7
	$\bar{X}$	+ 0.25	- 0.30	- 0.67	- 0.25
Low Neuroticism	+2	3	0	0	
	+1	3	4	1	
	0	1	2	4	
	-1	1	1	3	
	-2	0	1	1	
	-3	0	0	0	
	N	8	8	9	25
	ΣX	+ 8	+ 1	- 4	+ 5
	$\bar{X}$	+ 1.00	+ 0.42	- 0.44	+ 0.20
Totals	N	26	26	26	78
	$\Sigma \underline{X}$	+ 3	-11	-22	-30
	X	+ 0.12	- 0.42	- 0.85	- 0.38

cell-positions in the reduced table as those of the original Table 15.

Table 16

SUMS OF SQUARES AND VARIANCES OF NINE SUB-POPULATIONS CLASSIFIED BY MPI-N AND ACE

		Low ACE	Medium ACE	High ACE
High MPI-N	ΣX^2	15	13	32
	s^2	1.12	0.41	2.00
Medium MPI-N	ΣX^2	10	15	26
	s^2	1.36	1.57	2.75
Low MPI-N	ΣX^2	16	9	8
	s^2	1.14	1.27	0.77

Table 16 shows the largest variance, 2.75, is located in Cell_{23}, the smallest, 0.41, is located in Cell_{12}. Dividing the larger by the smaller yields the statistic F_{max} = 6.71. For 9 mean squares, and the greatest number of degrees of freedom = 9, tables of F_{max} show that there are 5 chances in 100 of obtaining a value as large as 9.45, 1 chance in 100 for a value of 14.7. With alpha = .05, one may confidently conclude that an obtained F_{max} = 6.71 supports the null hypothesis which dictates that these nine obtained variances do not significantly depart from homogeneity.

Normality of distribution of each of the nine cells is tested by applying the Kolmogorov-Smirnof one-sample

goodness-of-fit test to each of the nine distributions individually. This test, which assumes measurement on at least an ordinal scale and which requires that the criterion measure be continuous rather than discrete, is suited to the observed data because of the large proportion of small expected frequencies among the distributions. Table 17 has been devised to include all tests in one table, thus avoiding the repetition of information in the heading. Nevertheless, it is a composite of nine tables, each containing the same class of information as Table 6, where one distribution was tested for normality.

In Table 17, expected cumulative proportions were determined directly from tables of the unit-normal probability curve, dividing the baseline into six equal units each representing one standard deviation from the mean (the small residue beyond plus and minus three standard deviations was included in the outer categories). Observed cumulative proportions were based on the N of each distribution. Subtracting the one from the other, without regard for sign, indicated which D-score category contained the greatest deviation from normal expectation, or the statistic D.

Table 17 shows that eight of the nine distributions depart from normality to a degree that does not approach the significance level of .05. One, in Cell_{31}, departs drastically ($P < .01$).

It was earlier noted that a significant F-ratio may be attributed to an unwarranted null hypothesis when all other

Table 17

GOODNESS OF FIT OF NORMAL DISTRIBUTION WITH EACH OF NINE SUB-CLASS DISTRIBUTIONS CLASSIFIED ON MPI-N AND ON ACE*

D-Score Category:	-3	-2	-1	0	+1	+2	P -
Expected Cum. Prop.:	.0228	.1578	.5000	.8413	.9772	1.0000	Value
Cell$_{11}$ (N=10) f:	0	3	2	4	1	0	
Observed Cum. Prop:	.0000	.3000	*.5000	.9000	1.0000	1.0000	
Difference (O-E):	.0228	(.1422)*	.0000	.0587	.1578	.0000	$P > .20$
Cell$_{12}$ (N=8) f:	0	2	5	1	0	0	
Observed Cum. Prop.:	.0000	.2500	.8750	1.0000	1.0000	1.0000	
Difference (o-E):	.0228	.0922	(.3750)	.1578	.0228	.0000	$P > .15$
Cell$_{13}$ (N=8) f:	2	3	1	1	1	0	
Observed Cum. Prop.:	.2500	.6250	.7500	.8750	1.0000	1.0000	
Difference (O-E):	.2272	(.3672)	.2500	.0337	.0228	.0000	$P > .15$
Cell$_{21}$ (N=8) f:	0	1	0	4	2	1	
Observed Cum. Prop.:	.0000	.1250	.1250	.6250	.8750	1.0000	
Difference (O-E):	.0228	.0328	(.3750)	.2163	.1022	.0000	$P > .15$
Cell$_{22}$ (N=10) f:	0	2	3	1	4	0	
Observed Cum. Prop.:	.0000	.20000	.5000	.6000	1.0000	1.0000	
Difference (O-E):	.0228	.0422	.0000	(.2413)	.0228	.0000	$P > .20$
Cell$_{23}$ (N=9) f:	2	1	1	2	3	0	
Observed Cum. Prop.:	.2222	.3333	.4444	.6667	1.0000	1.0000	
Difference (O-E):	(.1994)	.1755	.0445	.1746	.0228	.0000	$P > .20$
Cell$_{31}$ (N=8) f:	0	0	1	1	3	3	
Observed Cum. Prop.:	.0000	.0000	.1250	.2500	.6250	1.0000	
Difference (O-E):	.0228	.1578	.3750	(.5913)	.3522	.0000	$P < .01$
Cell$_{32}$ (N=8) f:	0	1	1	2	4	0	
Observed Cum. Prop.:	.0000	.1250	.2500	.5000	1.0000	1.0000	
Difference (O-E):	.0228	.0328	.2500	(.3413)	.0228	.0000	$P > .20$
Cell$_{33}$ (N=9) f:	0	1	3	4	1	0	
Observed Cum. Prop.:	.0000	.1111	.4444	.8889	1.0000	1.0000	
Difference (O-E):	.0228	(.0467)	.0445	.0466	.0228	.0000	$P > .20$

* after Siegel (6, p. 47) and Tables of D which do not show values of D for $P > .20$.

** the greatest difference (statistic D)is designated by parentheses.

assumptions can be made. In the present instance there are two possibilities. First, one may assume that Cell_{31} is normally distributed in the parametric population. Second, one may conjecture whether a significant F-ratio based on an analysis of variance of Table 15 can be attributed to an unwarranted null hypothesis or to the departure of Cell_{31} from normality. The answer is easily ascertained but, for illustrative purposes, it will be set aside until the analysis of variance has been performed. Basically, the answer can be found by constructing a hypothetical distribution of D-scores which is identical to the obtained distribution in every respect excepting that a slight alteration in Cell_{31} will be effected so that this cell does not significantly depart from normality. With this hypothetical distribution, a completely new analysis of variance will be performed to determine whether there is any difference between an F-ratio based on an obtained distribution having one cell departing from normality, and a hypothetical distribution in which all assumptions are met. If there is none, then the significance of the obtained F-ratio it may be concluded, cannot be attributed to the errant Cell_{31}.

In addition to the usual assumptions, the computational procedure for a 3x3 factorial design usually assumes equal numbers of subjects in each cell. This had been the original intent but, because of unavoidable loss of subjects, cell entries are now slightly different from the 8.667 subjects that equality would require in each cell. There are, in fact, five cells having N = 8, two having N = 9 and two having N = 10.

That is, five cells depart from equality by two-thirds of a subject each, two depart by one-third each, and two by one and one-third subjects each. This departure is probably insufficient to contraindicate the use of the F-test. Indeed, the general concensus of statisticians is that, even in the face of slightly broken assumptions, the F-test remains a powerful instrument for testing the null hypotheses for which it was designed.[1] Computation procedures used in the present investigation will remain those drawn directly from the mathematical model of Table 14, the only difference being the necessity of dividing each sum of squares by its appropriate N rather than by dividing the summed sum of squares by their common N. The two procedures are arithmetically identical.

Notwithstanding the foregoing, some doubt may linger with respect to the effects of unequal cell entries. To allay these doubts, an additional analysis of variance will later be produced to show that there is no significant difference between F-ratios obtained on the above procedure and those derived from a procedure designed to yield approximate F-ratios from unequal cell entries based on means. This approximate procedure is normally employed when cell entries differ much more than they do in Table 15. This analysis will not be presented until the one with which it is to be compared is presented.

---

[1] A recent comment of Edwards (2, p. 132) is pertinent to this point. He states: "The F test of analysis of variance in other words, remains a robust test under a variety of violations of the assumptions on which it is mathematically based."

Testing Hypotheses Four, Five and Seven

Earlier in this chapter it was stated that tests of Hypotheses Four and Five, in Chapter Fourteen, must be regarded as tentative in so far as the variables were individually and not independently investigated. The two-way design of the present section permits independent testing of H_r, H_c and H_{rc}, the null hypotheses drawn from the mathematical model, in a single analysis. The pertinent Basic Hypotheses are repeated here:

Four: That increasing the external stress stimulus through ego-threatening or failure stress will depress the effectiveness of high-neuroticism subjects and will enhance the effectiveness of low-neuroticism subjects.

Five: That increasing the external stress stimulus through ego-threatening or failure stress will depress the effectiveness of subjects evidencing experience of high scholastic aptitude and will enhance the effectiveness of subjects evidencing experience of low scholastic aptitude.

Seven: That the three personality variables of Hypotheses 4, 5 and 6 vary independently, both among themselves and among their effects upon performance; that is, their effects are additive.

The null hypothesis H_4 simply states that the row means of Table 15 (right-hand marginal entries) are equal. The null hypothesis H_5 similarly states that the column means of Table 15 (lower marginal entries) are equal. It will be recognized that, because of the implications of $F < 1$ discussed in Chapter Fifteen, Basic Hypothesis Seven now applies only to the two personality variables of Hypotheses Four and Five. Thus H_7 is the null hypothesis that interaction equals zero. These three hypotheses are tested by the F-ratios shown in Table 18.

Table 18

ANALYSIS OF VARIANCE OF D-SCORES SHOWN IN TABLE 15

Source of Variation	Sum of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Row Means: Neuroticism	21.42	2	10.71	7.71	3.13	4.92
Column Means: Sch. Aptitude	12.07	2	6.04	4.35	3.13	4.92
Rows x Columns: Interaction	3.30	4	0.83	< 1.00	2.50	3.60
Error: Within Cells	95.67	69	1.39			

The confidence coefficient for this study was set at .95, that is alpha = .05. The probability of an obtained F = 7.71 for row means under H_4 is approximately P = .001; of the obtained F = 4.35 for column means under H_5 is approximately P = .035. The interaction F-ratio of less than unity indicates a very strong probability that H_7 is correct.

The probability values under H_4 and H_5 are strikingly similar to those obtained when H_4 and H_5 were tested with respective one-way analyses. This observation indicates that the variables actually were independent in their effects when tested previously, a conclusion which is now supported by the finding that interaction is near-zero (H_7).

Before proceeding to draw inferences from these data it is appropriate to investigate the effects of the earlier finding that certain assumptions were not rigorously met.

Effect of Non-Normal Cell$_{31}$

Under the hypothesis that the obtained significant F-ratios of Table 18 cannot be attributed to the effect of the errant Cell$_{31}$, a hypothetical set of data was constructed that was identical to the data in Table 15 in every respect except one: a single D-score in Cell$_{31}$ was altered from a value of +1 to 0. Table 19 tests the goodness of fit of this hypothetical Distribution$_{31}$ with a normal distribution in the manner that has been used previously. The greatest departure from normal expectation is shown by the proportion .4663 which, by reference to Tables of D, would occur in slightly more than 5 trials out of 100 under the null hypothesis. It is concluded that the hypothetical distribution of Cell$_{31}$ does not significantly depart from normal.

Table 19

GOODNESS OF FIT OF HYPOTHETICAL DISTRIBUTION$_{31}$
WITH A NORMAL DISTRIBUTION, $N_{31} = 8$

D-Score:	- 3	- 2	- 1	0	+ 1	+ 2
Expected Proportion:	.0228	.1578	.5000	.8413	.9772	1.0000
Observed (Hyp.) f:	0	0	1	2	2	3
Observed Proportion:	.0000	.0000	.1250	.3750	.6250	1.0000
Difference (O-E):	.0228	.1578	.3750	(.4663)	.3522	.0000

With the assumption of normality now met, the data were recalculated to produce the analysis of variance shown in Table 20.

Table 20

ANALYSIS OF VARIANCE OF HYPOTHETICAL DISTRIBUTION OF D-SCORES (ONE SCORE ALTERED FROM TABLE 15)

Source	S.S.	df	M.S.	F	$F_{.95}$	$F_{.99}$
Rows	20.28	2	10.14	7.24	3.13	4.92
Columns	11.14	2	5.57	3.98	3.13	4.92
Interaction	2.71	4	0.68	< 1.00	2.50	3.60
Error	96.55	69	1.40			

There is very little difference between the F-ratios of Tables 20 and 18. As nearly as can be estimated, the probability values have changes as follows: for rows, .001 in Table 18 and .002 in Table 20; for columns, .035 in Table 18 and .04 in Table 20; for interaction both are well within the range of acceptance. It is concluded, therefore, that the hypothesis of no difference between the obtained distribution of 78 D-scores and the hypothetical distribution in which one score is altered, is upheld and that the inference that the analysis of variance in Table 18 is securely based is justified.

Effect of Unequal Cell Entries

An approximate method for analyzing the variance of a two-way classification when cell entries are unequal is provided by Walker and Lev (8, pp. 381-382) to simplify the more complex computational procedures that otherwise would be necessary. The computational procedures employed in arriving at the sums of squares of Table 18, although complex, were the arithmetical equivalent of the standard procedure. In the unlikely event that an inflation of F-ratios is suggested to be a result of unequal N's, the findings of an analysis according to the Walker and Lev approximation will be found interesting. This method is based almost entirely upon mean cell values, treating each as a single entry, and uses individual scores only to arrive at the mean square for within cells which is corrected by using the mean reciprocal of cell N's as a multiplier. The analysis is presented in Table 21.

The most striking feature of Table 21 is that although the F-ratio for interaction is still much less than unity, the F-ratios for rows and columns are very large, that for columns now going beyond the .01 level of significance and that for rows much beyond the .001 level. These are statistically much more significant than those of Table 18, and each of them shows $P < .05$. It must be cautioned that the findings reported in Table 21 are approximate inasmuch as means were used as cell entries. Nevertheless, if confirmation for the significance of obtained F-ratios in Table 18 is sought, Table 21 provides it.

Table 21

ANALYSIS OF VARIANCE OF D-SCORES SHOWN IN TABLE 15, CALCULATED BY THE METHOD OF UNEQUAL FREQUENCIES IN SUBCLASSES[1]

Source	S.S.	df	M.S.	F	$F_{.95}$	$F_{.99}$
Rows	2.7734	2	1.3867	8.58	3.13	4.92
Columns	1.6805	2	0.8403	5.10	3.13	4.92
Interaction	0.1205	4	0.0301	<1.00	2.50	3.60
Error	----	69	0.1617			

[1] Walker and Lev (3, pp 381-382)

Anxiety and Neuroticism

The design of the basic investigation favoured Eysenck's operational definition of neuroticism (MPI-N) over Taylor's operational definition of anxiety (MAS) because (a) it seemed more closely related to the variable under study and (b) the MAS is significantly correlated with extraversion - introversion which is not under study. Because of the widespread interest in the anxiety variable as defined by Taylor (for example, see entries in the *Journal of Experimental Psychology* of the early and middle 1950's), it would be interesting to learn the differential effects on the high-anxious, medium-anxious and low-anxious subjects of the stress-inducing supervision used in this experiment. For this purpose, the MAS was administered along with the other personality measures prior to the experiment.

The method of comparison was the same as in Table 15, excepting that MAS levels were used as a basis of classifying subjects by columns (instead of ACE). The column means were in the same direction as the MPI-N row means, but the differences were not so great. Low-anxious subjects (N=25) had a mean D-score showing an improvement of + 0.08. Medium-anxious subjects (N=27) had a D-score mean of - 0.44 (very close to the general average - 0.38), and high-anxious subjects deteriorated an average of - 0.77. The corresponding means for low, medium, and high neuroticism subjects, from Table 15, are + 0.20, - 0.25, and -1.08.

The variance of the D-scores classified on the bases of these two variables, neuroticism and anxiety, was analyzed in the same way that has been done elsewhere in this chapter, with the same values of $F_{.95}$ and $F_{.99}$. The F-ratio for neuroticism means, of course, was significant ($P < .05$) as before, with a probability of approximately $_{.001}$. The F-ratio for column means of subjects classified by level of anxiety (MAS) of 3.16 indicates that the null hypothesis of no differences must be rejected when the investigator is working at the .05 level ($F_{.95} = 3.13$). Interaction was not significantly different from zero.

One may safely conclude, therefore, that the personality characteristic or characteristics known as anxiety, neuroticism, or neurotic anxiety has an effect on determining a student teacher's change in general teaching behavior when submitted to stress-inducing supervision, but that this effect

is most clearly demonstrated by scores on the N-variable of the MPI. It is possible, however, that were the anxiety variable measured with an instrument that was free from the effects of introversion-extraversion (see following section), it would operate in the same way as neuroticism. It is interesting to note that the correlation between the MAS and MPI-N, often quoted as .77, was corroborated in the present investigation where it was found to be .775.

Extraversion-Introversion

In view of the fact that both Thompson and Hunnicutt (7) and Forlano and Axelrod (4) found extraversion-introversion to be a powerful determinant of behavior under conditions of praise and blame (praise: assigning evaluations of success; blame: assigning evaluations of failure) one might expect this variable to operate the same way with respect to student teachers under conditions of failure-stress. Thompson and Hunnicutt found, for example, that (a) praised introverts achieved better than blamed introverts, (b) blamed extraverts achieved better than praised extraverts, (c) blamed extraverts and praised introverts did equally well, and (d) praised extraverts and blamed introverts did equally poorly. This study was reported in 1944 and replicates, under better design and conditions, the other study which, reported in 1937, produced the same results.

It was possible to investigate this relationship in the present investigation because Eysenck's MPI contains a measure of the extraversion-introversion variable. This was

not, however, a part of the basic investigation, nor did it bear any relationship to the basic hypotheses being tested. It is reported here partly because of general interest in possible personal variables that may affect the criterion, and partly because it may help to explain the difference between effects of anxiety (MAS) and neuroticism (MPI-N). The D-scores of the 78 subjects were classified according to high MPI-E scores (highly extraverted), medium MPI-E scores (in-between), and low MPI-E scores (highly introverted). The mean D-scores of subjects so classified were - 0.32, - 0.44 and - 0.46, respectively. It would not be expected that means so similar as this could, upon statistical analysis, result in an F-ratio significant at the .05 level and, of course, they did not. The F-ratio of variance between means to variance within groups was less than unity. Evidently the personality variable of extraversion-introversion, as Eysenck defines it ("extraversion, as opposed to introversion, refers to the outgoing, uninhibited, social proclivities of a person,") does not differentiate the performance of student teachers submitted to failure-stress. Either this conflicts with the studies of the 1930's and 1940's or these studies employed an operational definition of the variable that is different from Eysenck's. That the latter could easily be so is inferred (a) from inspection of their instrument, Pintner's Aspects of Personality, the "introversion-extroversion" items of which appear to be measuring something other than the variable defined above, and (b) from the fact that Cronbach (1)

in reporting extensively on one of these studies, makes no mention of extraversion-introversion but, rather, refers to the two groups as the (1) retiring, worried, sensitive and withdrawing, and the (2) self-assured, outgoing and active. One cannot help but wonder whether Pintner's measure is not, in fact, heavily loaded with Eysenck's neuroticism variable so that these earlier studies, reinterpreted, actually support the findings of the present investigation.

Conclusions

This chapter was designed to provide information with which to test three null hypotheses: H_4 that there is no difference between mean D-scores for subjects classified by level of neuroticism when this variable is investigated independently of the effects of scholastic aptitude; H_5 that there is no difference between D-scores of the same subjects classified by level of scholastic aptitude when this variable is investigated independently of the effects of neuroticism; and H_7 that interaction is zero, the two variables being independent in their main effects and the cell means being proportional to marginal means.

The information provided in Table 18, confirmed and supported by Tables 20 and 21, rejects H_4 ($P < .05$, approximately .001), and H_5 ($P < .05$, approximately .035), and supports H_7 ($P > .05$). Referring to the hypotheses which were deduced from the theory in Chapter Eight, it may now be concluded that this investigation has provided research evidence in support

of Basic Hypotheses Four, Five and Seven (as amended to exclude the need-for-achievement variable). Evidently, in the assessment of the fifteen experimenters, the effectiveness of a student teacher in a classroom situation does change when the supervisory figure induces failure stress in the subject. This is, however, a differential effect: those student teachers who are most neurotic and intelligent exhibit a deterioration in their general teaching effectiveness. Those who are least neurotic and (for this group) least intelligent improve. The general tendency for the group of 78 student teachers, as well as for those who fell in the mid-range on each of these two personality variables, was toward a slight deterioration in their general teaching behavior. The effects attributable to neuroticism (MPI-N) are the same as those attributable to anxiety (MAS) although the latter variable does not differentiate the subjects to nearly the same degree of statistical significance as the former, probably the result of the correlation of .35 existing between the MAS and a measure of extraversion (MPI-E) which does not differentiate the subjects at all.

REFERENCES FOR CHAPTER XV

(1) Cronbach, Lee J. Educational Psychology. New York: Harcourt Brace, 1954.

(2) Edwards, Allan L. Experimental Design in Psychological Research (Revised Ed.). New York: Rinehart, 1950.

(3) Ferguson, G. A. Statistical Analysis in Psychology and Education. New York: McGraw-Hill, 1959.

(4) Forlano, G. and H. C. Axelrod "The Effects of Repeated Praise or Blame on the Performance of Introverts and Extroverts," Journal of Educational Psychology, 28 (1937) 92-100.

(5) Lindquist, E. G. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton-Mifflin, 1953.

(6) Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

(7) Thompson, G. G. and C. W. Hunnicutt "The Effect of Repeated Praise or Blame on the Work Achievement of 'Introverts' and 'Extroverts'," Journal of Educational Psychology, 35 (1944) 257-266.

(8) Walker, H. M. and J. Lev Statistical Inference. New York: Henry Holt, 1953.

CHAPTER XVI

TREND ANALYSIS OF THE EFFECTS OF COMBINED DRIVE AROUSIBILITY INTENSITIES

The theoretical orientation in Chapter Eight posited at least three components of the internal dimension: neuroticism, need for achievent, and scholastic aptitude. It was suggested that there may be more. These component elements of personality were presumed to combine in some way to form the drive arousibility level of an individual which, together with factors of the external world, constitute the individual's drive potential for a given situation. A marked increase in objective drive stimuli was hypothesized to result in the differential effects upon performance that have, in Chapter Sixteen, been found to depend upon two of the internal or personality factors.

Combination of Main Effects

The dependency of performance upon each of the two main variables under study was established, however, only when the factors were studied independently. The theory, on the other hand, hypothesized that the differential effects would also depend upon their combination and, in the absence of indications otherwise, upon their additive combinations. Eventually, in Chapter Nine, this became phrased as Basic Hypothesis Eight:

> That, taking Hypotheses 1 to 7 together, the direction of change in teaching effectiveness of subjects submitted to stress-inducing supervision is predictable from a combination of the three personality variables stated therein: the higher the combination the greater the liklihood of a performance decrement, the lower the combination the greater the liklihood of a performance increment.

The hypothesis is qualified by the hypothesized mean decrement (supported earlier) which would tend to depress all scores, and is further qualified by the necessity of restricting it to two variables for purposes of testing.

The data provided in Table 15 are appropriate to test Hypothesis Eight. This is so because of the nonsignificant and near-zero interaction found in the table. It is now possible to combine the two variables so as to yield several levels of drive arousibility intensity, and to test Hypothesis Eight with an analysis of variance of D-scores so classified.

Such a classification will also permit a test of Hypothesis Nine:

> That there will be a greater degree of performance change (incremental or decremental) shown by subjects who are at the upper and lower extremes of the personality variables than by subjects who are in the middle range.

This hypothesis was necessary in order to accommodate the theoretical concept of a curvilinear relationship between effective drive potentials and effective reaction potentials. The nature of the curve itself will not be tested but, rather, just the essential implication: that performance change is greater at the outer levels of drive intensity than at the inner levels.

Five Levels of Intensity

The nature of the classification proposed in this chapter is evident from Table 22. This table uses the information taken directly from Table 15, retaining the same 3x3 layout and cell-locations. In Table 22, however, the only information entered into the cells is the respective number of subjects (N), mean D-score ($\overline{X}$), and a combination of two capital letters from H, M and L representing High, Medium and Low classification of this cell on each of the two marginal dimensions. In this manner, cell$_{11}$ contains the entry HL denoting its contents as high MPI-N, low ACE subjects, and so forth to Cell$_{33}$ which is designated LH representing its subjects as possessing low MPI-N and high ACE scores.

Table 22

TWO-WAY CLASSIFICATION OF MEAN D-SCORES SHOWING FIVE LEVELS OF INTENSITY OF TREATMENT-COMBINATIONS

Level of Neuroticism (MPI-N)	Level of Scholastic Aptitude (ACE)		
	Low (L)	Medium (M)	High (H)
High MPI-N (H)	-0.700 HL (N=10)	-1.125 HM (N=8)	-1.500 HH (N=8)
Medium MPI-N (M)	+0.250 ML (N=8)	-0.300 MM (N=10)	-0.667 MH (N=9)
Low MPI-N (L)	+1.000 LL (N=8)	+0.125 LM (N=8)	-0.444 LH (N=9)

Reading down the diagonals from upper left to lower right reveals a homogeneous classification of subjects by treatment-combinations. If order of cell-designations (i.e. MH or HM) is disregarded, and if the relative weight of each factor is ignored (both were found to be significant, but ACE less so than MPI-N), then these five treatment-combinations may be regarded as five levels of intensity of drive arousibility. The five levels are as follows:

5.	High	HH
4.	Medium High	MH
3.	Medium	MM or LH
2.	Medium Low	ML
1.	Low	LL

It is assumed that a low score on one measure and a high score on the other tend to combine to result in some middle-level (Medium) of intensity. This was suggested in the chapter on theoretical orientation.

Analysis of Diagonals

Referring back to the general Table 15 for the actual D-scores of subjects classified by diagonals, it is possible to calculate for each intensity-classification the mean, sums of squares and variance. For illustrative purposes, Table 23 provides the N, mean, and deviation from the general mean, of each group of subjects classified by the diagonals of treatment-combinations.

Table 23

MEAN D-SCORES OF 78 STUDENT TEACHERS CLASSIFIED BY INTENSITY OF DRIVE AROUSIBILITY FACTORS (MPI-N, ACE)

Intensity of Drive Arousibility	N	Mean D-Score	Deviation from General Mean = -0.38
High	8	- 1.50	- 1.12
Medium High	17	- 0.88	- .50
Medium	29	-0.48	- .10
Medium Low	16	+ 0.19	+ .57
Low	8	+ 1.00	+ 1.38

The data in Table 23 reveal not only mean differences between the classification levels, but also a mean trend. It would appear to be the trend that was suggested in Hypotheses Eight and Nine: that the higher the combination the greater will be the performance decrement, the lower the combination the greater will be the performance increment, and that the size of increment or decrement depends upon the degree to which the classification level departs from the mean. Were it not for Hypothesis Nine, one might suppose the difference depended only on the direction from the mean, all those on one side improving to the same extent, and all those on the other deteriorating equally.

When the presence of a trend is suspected or hypothesized, the data may be submitted to a trend analysis. This

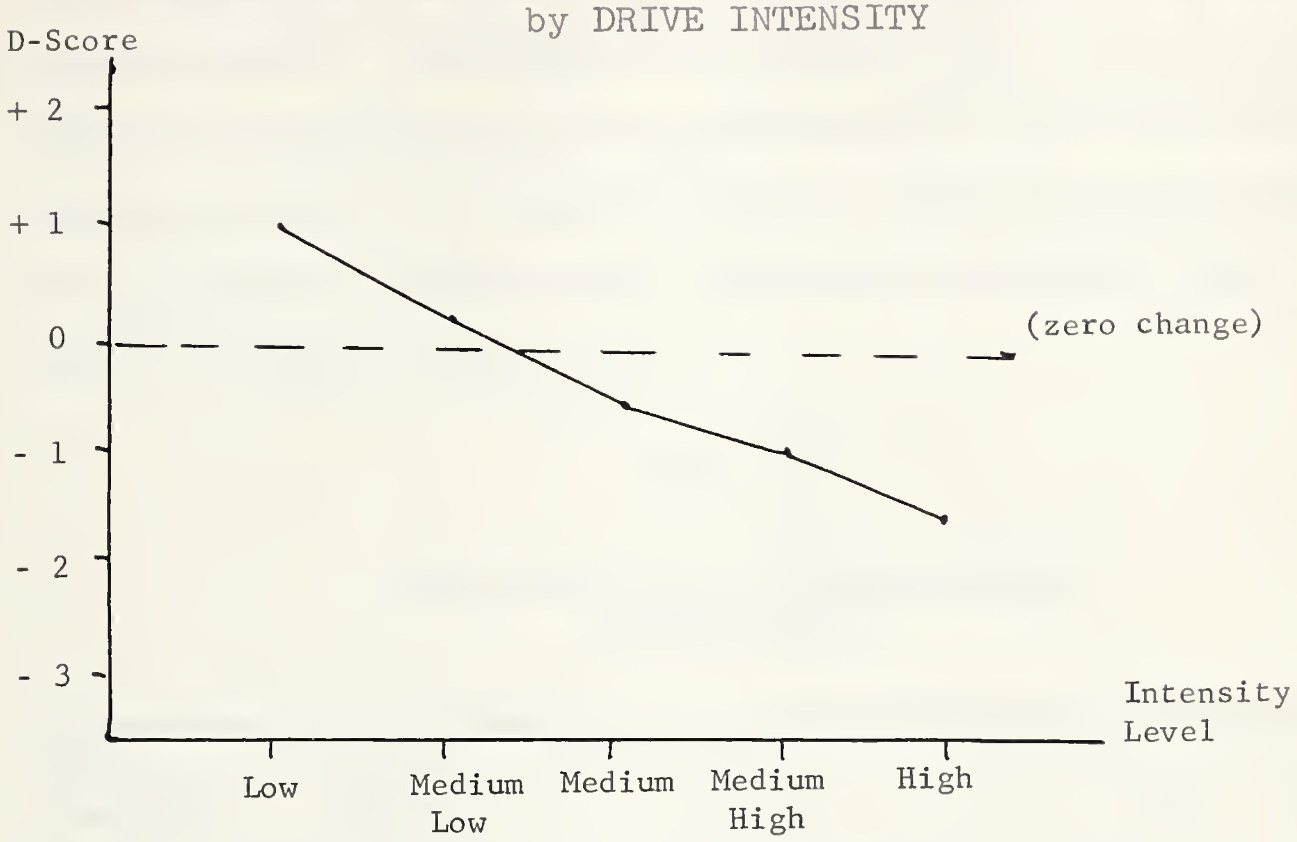
is an analysis of variance designed, in the first stage, simply to test the presence of a trend and, in the second stage, to test the significance of the particular trend. A trend analysis is particularly appropriate to the present data which has already been arranged in a general variance design, the assumptions of which have been tested. Equality of intervals between classification levels is not assumed for the first stage but is for the second. To establish the presence of a trend requires repartitioning the sums of squares derived from the general Table 15 with attention being given to the diagonals rather than to rows and columns. The two variance estimates, after repartitioning, become the variance for (1) between diagonal means and for (2) within subclasses. This is the simple one-way analysis of variance in which each variance is an estimate of the same population variance. The second stage of trend analysis divides the variance for between means into two components: linearity and deviations from linearity.

As a first step, the means are plotted against D-score categories, as in Figure 5, to observe the trend from one level of intensity to another. Figure 5 shows that the tendency of means is to decrease in value as the intensity value increases. This decreasing monotonic sequence is the relationship that was hypothesized in Basic Hypotheses Eight and Nine and which will now be tested.

Figure 5 represents the Mean- D-scores of Table 23 in diagram form showing an almost-straight line from a mean

Figure 5

DEGREE AND DIRECTION OF CHANGE IN PERFORMANCE UNDER STRESS, EXPRESSED AS A TREND IN MEANS OF SUBJECTS CLASSIFIED by DRIVE INTENSITY



gain of +1 for low-arousibility subjects to a mean decrement of -1.5 for high-arousibility subjects or, rather, for subjects in whom there is respectively a low and high combination of the two drive-arousibility factors investigated earlier: neuroticism and scholastic aptitude.

The question posed by Figure 5 would read: Is this really a trend or is the difference between this line and a straight horizontal line but a chance fluctuation? This leads to the null hypothesis that states (H_8) that there is no difference between mean D-scores and, hence, an absence of

trend. The test is the F-ratio MS_b/MS_w, the mean square for within groups being calculated from the five treatment groups by the Walker and Lev method for unequal frequencies in a one-way layout. This method was earlier used in the one-way analysis of variance of scores classified on the basis of neuroticism only, in Chapter Fourteen, and is an exact method. The analysis of variance data required to test the trend hypothesis is shown in Table 24.

Table 24

TRENDS ANALYSIS OF DIAGONAL CELLS FROM TABLE 15

Source of Variation	Sum of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Total	132.46	77				
Between Means	35.02	4	8.76	6.59	2.50	3.60
Within Groups	97.44	73	1.33			

Table 24 shows an F-ratio of 6.59 which has a probability of occurance under the null hypothesis of less than .001. Therefore we may reject the null hypothesis H_8 and conclude with great confidence ($P < .05$) that the presence of a trend in mean D-scores of 78 student teachers classified according to their combined cell-designations on dimensions of neuroticism and scholastic aptitude cannot be attributed to chance fluctuations.

Significance of Linearity Component

Lindquist makes it clear (2, p. 99, pp 340 ff) that where the relationship between the predictor variable (in this instance, the levels of drive arousibility intensity) and the criterion variable (D-score) is expressed by a monotonic sequence, and where it is obviously so and could be so expected on *a priori* grounds, there is little need to investigate further. Nevertheless, investigators frequently wish to test the linearity of this trend, or to test its goodness of fit to any *a priori* trends. To do so, Lindquist recommends a procedure requiring the linear regression of one factor on the other on the basis of covariance analysis.

In quite a recent publication, however, Edwards (1, pp. 148 - 152 and pp. 224 - 253) gives very extensive treatment to trend analyses and relies heavily upon tables of coefficients for obtaining the linear and quadratic components of the treatment sum of squares, when treatments are equally spaced. These coefficients are simply "weights" by which are multiplied each of the subclass sums of scores and frequencies. The coefficients for linear components, for example, would, if plotted, form a straight line. The test for linearity becomes an F-test wherein the total sums of squares for treatment means is subdivided into the sums of squares for the linear component, and the sums of squares for deviations from the linear component. But the total sums of squares for the table, and the error variance (within subclasses) remains unchanged. In the present instance, Table 24 remains the same excepting that one

line -- between means -- becomes subdivided into two to indicate its components.

At this level of analysis, it is necessary that treatments be equally spaced; otherwise the coefficients would not form a straight line (e.g.: -2, -1, 0, 1 and 2 as Y-ordinates would, if X-increments were constant, fall on a straight line but would not if X-values varied). It is not difficult to make this assumption with the present data. Reference to Table 22 shows that the five new treatment levels were derived from the bases on which the data were classified in the first place, and that these were equally spaced. Scores on each variable were divided into three equal groups on the basis of a personality test and an intelligence test; the distribution on each test was approximately normal.

Edwards' procedure for analyzing trends is closely related to the procedures for making multiple-comparisons, or orthogonal comparisons within a treatment sum of squares. Hence the statistic D used in trend analysis means "difference" between sums negatively weighted and those positively weighted (or, weighted in whatever ways the hypothesis would dictate). Since the linear coefficients are arranged both positively and negatively about a zero point, all our obtained weighted sums of scores in the present investigation will be either positive or negative, depending on which end of the scale of D-scores (which are also positive and negative) we start from. This is as it is intended to be. The data necessary for an

analysis of the nature of the trend are arranged in Table 25; for convenience the letter a denotes the coefficient, the subscript refers to the subclass.

Table 25

ORTHOGONAL COEFFICIENTS AND OTHER DATA NECESSARY FOR TESTING THE SIGNIFICANCE OF LINEARITY BETWEEN 5 SUBCLASS MEANS

Subclass	Coefficients* (a_i)	a_i^2	N_i	$N_i a_i^2$	ΣX_i	$a_i \Sigma X_i$
H	- 2	4	8	32	- 12	+ 24
MH	- 1	1	17	17	- 15	+ 15
M	0	0	29	0	- 14	0
ML	+ 1	1	16	16	+ 3	+ 3
L	+ 2	4	8	32	+ 8	+ 16
Totals	0	10	78	97	- 30	D = 58

*from Edwards (1, Table XI, p. 377)

To find the sum of squares for the linear component, the value of the summed weighted sum of scores, or "D", is squared and divided by its appropriate degrees of freedom which equals the summed number of individuals in each subclass multiplied by the squared weighting. With equal entries in each subclass, the divisor is simply the number per subclass times the sum of the a_i^2 column. With unequal entries the same result is obtained by multiplying each N by its squared coefficient, and summing. From the data in Table 25, this process

arrives at the ratio $(58)^2/97$, or 34.68 as the sum of squares for the linear component with df = 1. The sum of squares for deviations from linearity is the difference between the sum of squares for between means (from Table 24, equals 35.02) and the linear component sum of squares, or 0.34. Each of these, as a component of the between means sum of squares, yields a mean square which is an estimate of the same variance and may be cast into the F-ratio of the larger over the smaller. This is done in Table 26.

Table 26 shows a very significant F-ratio of linear component variance to deviations from linear component variance; the value of 26.15 is well beyond $F_{.999}$. Working at the .05 level, therefore, the hypothesis that there is no difference between the variance that is attributable to linearity and to non-linearity is untenable. Not only may we conclude, from Table 24, that there is a definite and significant trend among the five subclass means, but that, from Table 26, this is a linear trend running from high (and positive) D-scores for subjects classified as low drive arousibility, to low (and negative) mean D-scores for subjects classified as high drive arousibility.

Conclusions

Chapter Sixteen concludes the analysis of Difference-Scores derived from the General Effectiveness criterion of the White Rating Sheets that were completed by the experimenters. The analyses provided support for Basic Hypotheses One, Two,

Table 26

TEST FOR SIGNIFICANCE OF LINEAR REGRESSION AND DEVIATIONS FROM LINEAR REGRESSION FOR THE DATA OF TABLE 25[1]

Source of Variation	Sum of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Linear Regression	34.68	1	34.68	26.15	3.98	7.01
Deviations from Linearity	0.34	3	0.11	< 1.00	3.98	7.01
Within Groups	97.44	73	1.33			
Total	132.46	77				

Three, Four, Five, Seven, Eight and Nine (when hypotheses seven to nine are amended to exclude the need for achievement variable of the non-supported Hypothesis Six). To this extent they indicated the value of the theory of interpersonal dynamics of supervisory relationships, contained in Chapter Eight, as

---

[1] Note: In applying Edwards' procedure to the present raw data there is necessarily a minor flaw. With unequal frequencies in the subclasses the $\Sigma N_i a_i$ should equal 0, just as it necessarily does with equal frequencies. In the present data, there are 17 MH individuals and 16 ML individuals so that $\Sigma N_i a_i = -1$. To show that this slight difference has no appreciable effect on the results, the data were recalculated after dropping one MH individual and reducing ΣX_{MH} by the $\bar{X}_{MH}$. This had the effect of making Linear Regression Sum of Squares = 33.84 and Sum of Squares for Deviations = 1.18. Their Mean Squares, respectively, of 33.84 and 0.39, yielded F values of 25.44 and < 1 which are infinitesimally different from those in Table 25. Alternatively, a very laborious procedure for arriving at appropriate coefficients to permit $\Sigma N_i a_i = 0$ could have been employed. It is doubtful that, in the present instance, any such modifications could affect the significance of the obtained F-ratio which is so far beyond the .001 level.

a valid set of concepts for interpreting different effects of stress-inducing supervision upon the general effectiveness of a student-teacher's behavior. The interested supervisor, however, also wishes to learn the effects of stress and tension upon particular aspects of the teaching activity. Chapter Seventeen presents one way of investigating this problem. It contains analyses of variance of D-scores derived from each of the component categories of the White Rating Sheet. Each analysis is conducted in essentially the same manner as that employed in Chapter Fifteen.

REFERENCES FOR CHAPTER XVI

(1) Edwards, A. L. Experimental Design in Psychological Research (Revised ed.). New York: Rinehart and Co., 1960.

(2) Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton-Mifflin, 1953.

CHAPTER XVII

THE COMPONENTS OF TEACHING EFFECTIVENESS

The preceding four chapters have provided information from which inferences regarding the behavior of student teachers subjected to failure stress could be drawn. These chapters referred only to the general teaching behavior of the subjects, the criterion was a comparison measure of teaching effectiveness that related lesson two to lesson one. The study has been diagnostic in the sense that differential behavior was traced to specific personality characteristics. The present chapter is somewhat more diagnostic in that the effects of failure stress are not only traced to personal attributes but also are examined with respect to each of several components of effective teaching: use of personal qualities, evidence of preparation for teaching, manner of presenting the lesson, contact with pupils, and classroom management. Now that the two factors of neuroticism and scholastic aptitude have been singled out as specific determinants of behavior change under stress, each of the components of teaching will be analyzed according to these two bases for classification of the appropriate D-scores.

These analyses, taken together, are the means of testing Basic Hypothesis Ten:

> That Hypotheses 1 to 9 inclusive, which apply to general effectiveness in the teaching task, will also apply to each of several component aspects of teaching effectiveness.

Personal Qualities

The student-teacher's personal qualities were defined on the White Rating Sheet as including poise, voice, manner, professional interest, and attitude. The D-score for personal qualities indicated the degree to which the subject tended to improve or deteriorate in the use of these attributes in the experimental lesson compared with the control lesson. The distribution of D-scores for the PQ (personal qualities) category is summarized in Table 27. This table is constructed in the same way as the general Table 15 excepting that individual scores are omitted; instead the sum of scores, number of individuals, and mean score for each of nine cells is shown. Because the same two bases for classification are used, the cell frequencies and marginal frequencies are the same as in Table 15. For convenience, the three values are shown in this form:

$$\frac{\Sigma X_{ij}}{N_{ij}} = \overline{X}_{ij}$$

in each cell.

Table 27 shows marginal means that tend to differ in the hypothesized direction. Personal qualities would appear to be sensitive to failure stress in much the same manner as was the general teaching criterion. The higher a subject's ACE score, the greater was the deterioration in personal qualities, the higher a subject's MPI-N score, the greater was the deterioration in personal qualities. There

Table 27

SUMS, FREQUENCIES AND MEANS OF 78 D-SCORES FOR PERSONAL QUALITIES, CLASSIFIED BY LEVEL OF NEUROTICISM (MPI-N) AND SCHOLASTIC APTITUDE (ACE)

Level of Neuroticism (MPI-N)	Level of Scholastic Aptitude (ACE)			
	LOW	MEDIUM	HIGH	Totals
HIGH	$\frac{-11}{10}$ = - 1.10	$\frac{-10}{8}$ = - 1.25	$\frac{-11}{8}$ = - 1.38	$\frac{-32}{26}$ = - 1.23
MEDIUM	$\frac{-2}{8}$ = = 0.25	$\frac{-10}{10}$ = - 1.00	$\frac{-8}{9}$ = - 0.89	$\frac{-20}{27}$ = - 0.76
LOW	$\frac{-1}{8}$ = - 0.13	$\frac{-5}{8}$ = - 0.63	$\frac{-9}{9}$ = - 1.00	$\frac{-15}{25}$ = - 0.60
Totals	$\frac{-14}{26}$ = - 0.54	$\frac{-25}{26}$ = - 0.92	$\frac{-28}{26}$ = - 1.08	$\frac{-67}{78}$ = - 0.86

are two important differences: (1) all the means are negative, not a single cell showed a mean improvement in personal qualities, and (2) the differences are not great. To test the significances the data are submitted to a two-way analysis

of variance similar to the method employed in Chapter Sixteen. Table 28 summarizes the analysis.

Table 28

ANALYSIS OF VARIANCE OF D-SCORES FOR PERSONAL QUALITIES FROM DATA IN TABLE 27

Source of Variation	S.S.	df	M.S.	F	$F_{.95}$	$F_{.99}$
Rows	5.64	2	2.82	2.66	3.13	4.92
Columns	4.18	2	2.09	1.97	3.13	4.92
Interaction	2.25	4	0.56	<1.00	2.50	3.60
Error	73.40	69	1.06			

The fact that not a single F-ratio approaches significance (.05) level is grounds for acceptance of all three null hypotheses: that there is no difference between means of rows (neuroticism), between means of columns (scholastic aptitude), and that interaction is not significantly greater than zero. Although Table 27 provided some very interesting information regarding the direction of change in personal qualities when a student teacher is subjected to failure stress, no inferences can be drawn from it in the face of non-significant F-ratios.

Preparation for Teaching

The degree to which a student teacher showed evidence of preparation for teaching was assessed by the experimenters in each lesson, experimental and control, and the direction

and degree of change was shown as the D-score for PT (preparation for teaching). The White Rating Sheets guided them to look for differences in general evidence of preparation, in academic knowledge, in the suitability of lesson materials, and use of effective planning. Table 29 summarizes the results of these assessments.

Table 29

SUMS, FREQUENCIES AND MEANS OF 78
D-SCORES FOR EVIDENCE OF PREPARATION FOR TEACHING,
CLASSIFIED BY LEVEL OF MPI-N AND ACE

Level of Neuroticism (MPI-N)	Level of Scholastic Aptitude (ACE)			
	LOW	MEDIUM	HIGH	Totals
HIGH	+1/10 = +0.10	+ 1/8 = + 0.13	+2/8 = +0.25	+4/26 = +0.15
MEDIUM	-2/8 = -0.25	+2/10 = +0.20	+3/9 = +0.33	+3/27 = +0.11
LOW	+6/8 = +0.75	+4/8 = +0.50	0/9 = 0.00	+10/25 = + 0.40
Totals	+5/26 = +0.19	+7/26 = +0.27	+5/26 = +0.19	+17/78 = + 0.22

It is difficult to make observations on the basis of the information in Table 29. It would seem, first of all, that the failure stress stimulus had an effect of sending student teachers back to their plans: nearly all cells showed an improvement in evidence of preparation for teaching. One

cell showed a slight mean decrement and another showed a zero mean. This observation would be in accord with earlier studies in anxiety which reported an increase in behavioral rigidity under stress. An interesting inference would be that where a supervisor sees strong evidence of preparation for teaching he may simply be witnessing a teacher acting under the stress his presence creates. But before inferences may be drawn, the statistical test is necessary. Table 20 summarizes the analysis of variance of D-scores for evidence of preparation for teaching, classified on the bases of two personality variables.

Table 30

ANALYSIS OF VARIANCE OF D-SCORES FOR PREPARATION FOR TEACHING FROM DATA IN TABLE 29

Source	S.S.	df	M.S.	F	$F_{.95}$	$F_{.99}$
Rows	1.24	2	0.62	1.02	3.13	4.92
Columns	0.10	2	0.05	<1.00	3.13	4.92
Interaction	4.08	4	1.02	1.67	2.50	3.60
Error	41.87	69	0.61			

Again, not a single F-ratio approaches significance (.05 level). The hypotheses, therefore, that row means are equal, that columns means are equal, and that interaction equals zero, must be accepted. Within the framework of the present theory and the hypotheses arising out of it, no valid

inferences may be drawn.

Presentation of Lessons

It was hypothesized (Basic Hypothesis Ten) that lesson presentation would suffer among those student teachers whose drive arousibility levels were high, and that it would be enhanced with those in whom they were low, when placed in the experimental situation. Evidence of improvement or deterioration in presentation of lessons (PL) was based on the subject's observed resourcefulness, effectiveness of methods, adaptability, and command of the English language. When an experimenter witnessed improvement along these lines he was directed to signify on the White Rating Sheet with a check mark in one of the positively-signed boxes, depending upon the degree of improvement in evidence. When deterioration was seen, the same procedure with respect to negatively-signed boxes was used. In a number of instances, no change was evident and a zero was checked. These D-scores for presentation of lesson were classified according to the subject's scores on the Eysenck Neuroticism scale and on the ACE Psychological Examination, in the same way that all other classifications had been made. When these were divided into the familiar nine cells representing three levels of each factor, the scores of each cell were summed and the mean taken. The sums, means and number of cell entries are shown in Table 30.

Table 31

SUMS, FREQUENCIES AND MEANS OF 78 D-SCORES FOR PRESENTATION OF LESSON, CLASSIFIED BY LEVEL OF MPI-N AND ACE

Level of Neuroticism (MPI-N)	Level of Scholastic Aptitude (ACE)			
	LOW	MEDIUM	HIGH	Totals
HIGH	-4/10 = - 0.40	-5/8 = -0.63	-8/8 = -0.44	-17/26 = - 0.65
MEDIUM	-3/8 = -0.38	-3/10 = -0.30	-4/9 = -0.44	-10/27 = - 0.37
LOW	+1/8 = +0.13	-5/8 = -0.63	-1/9 = -0.11	- 5/25 = - 0.20
Totals	-6/26 = -0.23	-13/26 = - 0.50	-13/26 = - 0.50	-32/78 = - 0.41

Marginal differences in Table 31 are slight. What trend there is would appear to be associated with the neuroticism factor where greater deterioration accompanies higher neuroticism. The only subpopulation of student teachers that did not exhibit a mean deterioration in their manner of presenting a lesson under failure stress was the one classified as low on both personal factors. From the experience of the preceding two analyses, it could be expected that the evidence in Table 31 would support the hypothesis that student-teacher group means for effectiveness of lesson presentation under stress do not significantly differ from one group to another, whether they are classified according to one factor or to the other. In fact, support for this null hypothesis is provided

by the analysis of variance in Table 32 in which none of the F-ratios approaches significance ($P > .05$).

Table 32

ANALYSIS OF VARIANCE OF 78 D-SCORES FOR LESSON PRESENTATION FROM DATA IN TABLE 31

Source of Variation	Sums of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Rows	2.69	2	1.35	1.32	3.13	4.92
Columns	1.25	2	0.63	<1.00	3.13	4.92
Interaction	2.83	4	0.71	<1.00	2.50	3.60
Error	70.10	69	1.02			

Contact with Pupils

One of the components of effective teaching on which student teachers are regularly evaluated in the student teaching program at the Faculty of Education, University of Alberta, is defined as contact with pupils (CP). This category, which is included on the White Rating Sheet as one of the criterion measures used in the present investigation, includes such necessary teaching behavior as showing sensitivity to the pupils' reactions during the lesson, adjusting the content and method to the grade level of the class being taught, and the manner the teacher uses for helping individual pupils with their difficulties. Evidently, CP does not imply physical contact but, rather, the interpersonal relations set up

between teacher and pupil to improve the learning situation. Experimenters attempted to evaluate this quality during each of the two lessons, the control and the experimental, and expressed any difference they noticed in the usual manner. It was hypothesized that this component, as all the others, would be affected by stress-inducing supervision differentially, according to the two predictor variables. The relevant statistics calculated from the obtained D-scores for PC appear in Table 33.

Table 33

SUMS, FREQUENCIES AND MEANS OF 78 D-SCORES FOR CONTACT WITH PUPILS, CLASSIFIED BY LEVEL OF MPI-N AND ACE

Level of Neuroticism (MPI-N)	Level of Scholastic Aptitude (ACE) LOW	MEDIUM	HIGH	Totals
HIGH	-4/10 = -0.40	-8/8 = -1.00	-11/8 = - 1.38	-23/26 = - 0.88
MEDIUM	-1/8 = -0.13	-7/10 = -0.70	-7/9 = -0.78	-15/27 = - 0.56
LOW	-1/8 = -0.13	-4/8 = -0.50	-7/9 = -0.78	-12/25 = - 0.48
Totals	-6/26 = -0.23	-19/26 = - 0.77	-25/26 = - 0.92	-50/78 = - 0.64

There are three pertinent observations than can be made with respect to Table 33. First, every cell mean is negative. Whatever the significance of the two predictors may be, it would seem that student teachers submitted to stress-inducing supervision do experience considerable difficulty in maintaining effective personal contact with their pupils. The implications of this observation are far-reaching: not only may a supervisory visit, if it causes tensions in a particular teacher, be expected to impair effective teacher-pupil relationships but also may one expect that a principal who effects an atmosphere of uncertainty and stress in his staff is actively curtailing the opportunity of the pupils to receive the full benefit of a teacher's personal resources.

A second observation is that the mean differences in Table 33 are larger than previous ones, suggesting that -- even though all subclasses deteriorated as groups -- certain teachers are more susceptible to ineffective pupil contact under stress. The third observation is that the differences are in the hypothesized direction, that contact with pupils tends to be impaired in the same way as general effectiveness although not enhanced in the same way so far as this group is concerned.

To put these last two observations to the test requires restating the null hypotheses that (1) there is no difference between means of subjects classified on the basis of neuroticism scores when investigated independently of

scholastic aptitude, (2) there is no difference between means of subjects classified by scholastic aptitude when the effect of the neuroticism variable is set aside, and (3) there is no interaction between the two. These hypotheses are tested with the analysis of variance, a summary of which is shown in Table 34.

Table 34

ANALYSIS OF VARIANCE FOR CONTACT WITH PUPILS, FROM DATA IN TABLE 33

Source of Variation	Sums of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Rows	2.39	2	1.20	1.10	3.13	4.92
Columns	7.26	2	3.63	3.33	3.13	4.92
Interaction	1.06	4	0.27	< 1.00	2.50	3.60
Error	75.24	69	1.09			

Table 34 gives strong support ($P > .05$) to the null hypothesis for interaction and leads, as well, to support ($P > .05$) for the null hypothesis for row means. Evidently, if stress-inducing supervision affects the teacher's contact with her pupils it does so without regard for her neurotic tendencies. However the obtained F-ratio for column means of 3.33 passes the .05 level of significance and leads to a rejection of the null hypothesis with respect to scholastic aptitude. It may, therefore, be concluded with confidence ($P < .05$) that among 78 student teachers in whom failure-stress

was induced there was a significant difference among their three subpopulations classified by level of scholastic aptitude, in the way they maintained contact with their pupils. Inasmuch as the manual to the A.C.E. Psychological Examination includes the term "general intelligence" as the function being measured, one must accept the conclusion that the more intelligent the student teacher, the more difficulty failure stress will cause to her contact with pupils, the less intelligent the student teacher the less will be the deleterious effect of stress.

Classroom Management

Because it is achieved in a variety of ways according to the individual teacher, classroom management is difficult to define. Some practice it, and evaluate it, in terms of rigid routine; others see it as synonymous with one use of the term 'discipline' and still others, claiming that effective management is a function of rapport and personal relations, refer to effectiveness in classroom management as a composite of all the good qualities in teaching. In the present investigation, the definition was that which is regularly used in the student teaching program. Classroom management (CM), then, refers to control of the class and the activities of individual pupils, efficient manner of giving directions and handling routine matters, and the manner of keeping records. To the extent that they were able, the experimenters noted the effectiveness of the student teacher's classroom management during the control lesson and compared it during the experimental lesson with what they observed during the first. D-scores for

the CM category were classified like the others; their results appear in Table 35.

Table 35

SUMS, FREQUENCIES AND MEANS OF 78 D-SCORES
FOR CLASSROOM MANAGEMENT, CLASSIFIED
BY LEVEL OF MPI-N AND ACE

Level of Neuroticism (MPI-N)	Level of Scholastic Aptitude (ACE)			
	LOW	MEDIUM	HIGH	Totals
HIGH	-2/10 = -0.20	-4/8 = -0.50	-8/8 = -1.00	-14/26 = - 0.54
MEDIUM	+2/8 = +0.25	-7/10 = -0.70	-9/9 = -1.00	-14/27 = - 0.52
LOW	+1/8 = +0/13	+2/8 = +0.25	-4/9 = -o.44	-1/25 = -0.04
Totals	+1/26 = +0.04	-9/26 = -0.37	-21/26 = - 0.81	-29/78 = - 0.37

A test of the extent to which observed differences between marginal means are statistically significant is provided by the F-ratios in Table 36. Of the three F-ratios in this two-way layout, only one is significant; it is the one which compares the variance attributable to column means (scholastic aptitude) with the variance within cells. Observed differences between mean D-scores of subjects classified by level of neuroticism are not statistically significant and may be attributed to chance fluctuations. The interaction is

not significantly different from zero. It is possible to conclude, however, that there is an observed difference between D-score means of subjects classified by level of scholastic aptitude that could occur under the null hypothesis in only approximately four instances out of 100. The more intelligent the subject, the more likely is the failure stress stimulus to depress her effectiveness of classroom management. The subjects who were in the lowest one-third of the experimental sample in terms of their ACE scores exhibited a slight mean improvement in their classroom management during the experimental lesson.

Table 36

ANALYSIS OF VARIANCE FOR CLASSROOM MANAGEMENT FROM DATA IN TABLE 35

Source of Variation	Sums of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Rows	4.02	2	2.01	1.66	3.13	4.92
Columns	9.34	2	4.67	3.86	3.13	4.92
Interactions	3.07	4	0.77	<1.00	2.50	3.60
Error	83.79	69	1.21			

Discussion of Findings

On first inspection this appears to be a most unusual situation: the global effectiveness ratings of degree and direction of change of behavior under stress were differentially affected according to two personality variables,

neuroticism and scholastic aptitude, but the similar ratings for each of five components of teaching effectiveness were not. Several conjectures give plausible explanations: (1) The five components do not exhaust the domain of aspects of effective teaching but are, instead, relatively peripheral aspects of the task. This does not seem defensible; these five, if not exhaustive, are commonly regarded as reasonably basic aspects of teaching. Accordingly, that which affects teaching _in toto_ should affect its parts. (2) The theory was wrong in predicting from these two variables; the findings of this chapter are definitive while those of the previous two chapters were the result of chance fluctuations. But this conjecture is too arbitrary in the face of probability values cited with respect to global effectiveness D-scores. (3) The D-scores for components were inadequately arrived at; those for global effectiveness were not. While this is only speculation, it is defensible. The experimenters had general experience and specific training for their task but were asked to perform the general rating first. The amount of time available for observing and rating was probably adequate for the global rating but insufficient to make the fine distinctions necessary to rate each of several components. The reliability of a test is increased by increasing its length; here there may have been insufficient length of observation time. It may be noted, too, that the analyses were performed on small groups,

approximately 26 subjects in each level of a variable. Increasing the size of the sample would have the effect of placing more confidence in mean differences that are observed. This is, in effect, what was done in the analysis reported in the following section where mean differences are based on "samples" of 78 subjects for each component.

Analysis of Variance Between Components

An attempt to determine the differential effects of two personality variables upon each of five components of the teaching activity when failure stress is induced revealed very little. Five interaction null hypotheses, in the preceding sections, were accepted. Eight of the ten null hypotheses with respect to mean differences were also accepted. The rejected two indicated that general intelligence was directly related to the difficulties that a teacher under stress-inducing supervision would experience in maintaining effective contact with her pupils and in managing the classroom.

The present section, and the next, were designed to reveal differences among the five component aspects of teaching in their sensitivity to psychological stress. The question to be answered would be: Quite apart from how well the predictors turned out with each component in turn, which of these five aspects of teaching were generally most affected by the experiment? It is an important question for supervisors. In Chapter Thirteen, the data revealed that the general or average tendency of subjects placed under stress-inducing supervision

is toward a slight deterioration in general teaching effectiveness. Now, the question is whether some parts of teaching tend to deteriorate in the same way as the general tendency, and which depart from this trend.

To test the hypothesis that there is no difference between mean D-scores for each of the component categories, the necessary data are arranged in Table 37. The order of the components differs from the order on the white rating sheets, and in discussions so far. Instead, they are arranged in order of mean magnitude from the largest positive mean to the largest negative mean.

Table 37

DATA FROM FIVE CRITERION MEASURES OF ASPECTS OF TEACHING EFFECTIVENESS, APPLIED TO EACH OF 78 STUDENT TEACHERS IN A FAILURE-STRESS EXPERIMENT

Measure:	PT	CM	PL	CP	PQ	Total
X	f	f	f	f	f	
+ 2	4	3	2	1	2	
+ 1	18	12	12	8	2	
0	50	32	27	30	23	
- 1	3	19	26	20	31	
- 2	3	8	11	17	14	
- 3	0	4	0	2	5	
ΣX	+ 17	- 29	-32	- 50	-67	- 161
N	78	78	78	78	78	390
$\overline{X}$	+ 0.22	- 0.37	- 0.41	- 0.64	- 0.86	
ΣX^2	51	111	90	118	143	513

Table 37 contains means ranging from + 0.22 for Preparation for Teaching to - 0.86 for Personal Qualities. It will be recalled that the General Effectiveness mean D-score was - 0.38, or approximately half-way between these two means. There are, of course, ten differences between the means provided in Table 37. To test the overall significance of differences between obtained mean D-scores for each of five components of effective teaching, the data are analyzed by the simple one-analysis of variance reported in Table 38. The highly significant F-ratio of variance between means to variance within groups ($P < .05$ and, in fact, $< .001$) renders untenable the null hypothesis of no differences. One cannot conclude, however, that all the observed differences in Table 37 are significant differences, only that there are significant differences among them.

Table 38

ANALYSIS OF VARIANCE OF DATA FROM TABLE 37

Source of Variation	Sums of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Total	446.54	389				
Between Means	50.76	4	12.69	12.32	2.39	3.36
Within Groups	395.78	385	1.03			

Multiple Range Comparisons of Component Means

To conclude only that the D-score means for the component categories differ significantly is not very satisfactory. The interested supervisor would want to know which aspects of teaching are more sensitive to stressful situations than which other aspects. The conventional method for determining which mean differences are significant employs the simple "Student's t" test (1, pp. 274 ff.). A more recent and somewhat more powerful test is Duncan's "New Multiple Range Test" (2, pp. 136-140), which is presented in this section.

The standard error of any mean[1] in Table 37, because of equal N's is

$$s_{\overline{X}} = \frac{s}{\sqrt{N}}$$

where "s", or standard deviation, is the square root of the error mean square of the analysis of variance (Table 38), and N is the number of observations on which the mean is based. Because "s" and "N" are constant for each distribution, $s_{\overline{X}}$ = 1,015/8.832 or .115. The "significant studentized ranges" for Duncan's new multiple range test are provided by Edwards. Each of these is a kind of "t" but their derivation recognizes that, from a number of differences, it seems reasonable to expect that we are more likely to have some real differences between the means than when only two means are compared.

---

[1] Because the calculations for s were originally corrected for sample size (N - 1), the whole $\overline{N}$ may be used in the denominator in this instance.

Although the investigation has, until now, utilized a confidence coefficient of .95, the choice of .99 for the present test is based on Duncan's concept of protection levels. A two-mean protection level is always given by $(1 - \alpha)$ which, if $\alpha = .01$, equals 99 per cent. If the null hypothesis is correct, then .01 is the probability that we will declare them to be different. But the k-mean protection level is lower and, for the multiple range test, is given as $(1 - \alpha)^{k-1}$. This is, again, because it is more likely that real differences will appear when $k > 2$ means are compared than when $k = 2$ means are compared. In our present investigation there are $k = 5$ means. Hence the protection level with $\alpha = .05$ is $(1 - .05)^{5-1}$ or $.95^4 = .81$, as the minimum probability of finding no erroneous significant differences among our five means. For this section, then, a high protection level will be maintained by working at the .01 level so that the protection level is $(1 - .01)^{5-1} = .99^4 = .96$ which is close to our original confidence coefficient of .95.

"Significant studentized ranges" at the .01 level for $k = 5$ means requires selecting those for each of four ranges: 2, 3, 4 and 5 representing the distances from one mean to another when means are arranged in order of magnitude. From these levels it is possible to calculate the "shortest significant ranges" in much the same manner as is done with conventional t-testing. These data are provided in Table 39.

Table 39

TABLE OF SIGNIFICANT STUDENTIZED RANGES (.01 LEVEL) AND SHORTEST SIGNIFICANT RANGES FOR k = 5 MEANS

Range	Significant Studentized Range (.01)	Standard Error of the Mean	Shortest Significant Range
k_2	3.970	0.115	.46
k_3	4.121	0.115	.47
k_4	4.225	0.115	.48
k_5	4.303	0.115	.49

From Table 39 it is evident that the greater the range in magnitude between any two means, the greater must be the observed difference to be regarded as significantly different from zero. The means of the present investigation, their differences, and the appropriate shortest possible range required to be significant, are presented in Table 40. Only those differences that are equal to or greater than the range shown to the right of the table are marked with an asterisk. For illustrative purposes, the letter-designation of each mean is reproduced directly below its column and "subsets" of means are underscored. Any mean within a subset is not significantly different from another mean within the same subset, but is significantly different from any mean not within this particular subset.

Table 40

DUNCAN'S NEW MULTIPLE RANGE TEST APPLIED TO THE DIFFERENCES BETWEEN k = 5 TREATMENT MEANS, FROM DATA IN TABLE 37

Category: Means:	PT + 0.22	CM - 0.37	PL - 0.41	CP - 0.64	PQ - 0.86	Shortest Significant Ranges
PT		.59*	.63*	.86*	1.08*	R_2 = 0.46
CM			.04	.27	.49*	R_3 = 0.47
PL				.23	.45	R_4 = 0.48
CP					.22	R_5 = 0.49
Subsets:	PT	CM	PL	CP	PQ	
			———	———	———	
		———	———	———		

Any two treatment means not underscored by the same line are significantly different (.01).

Any two treatment means underscored by the same line are not significantly different (.01).

---

* (significant at .01 level)

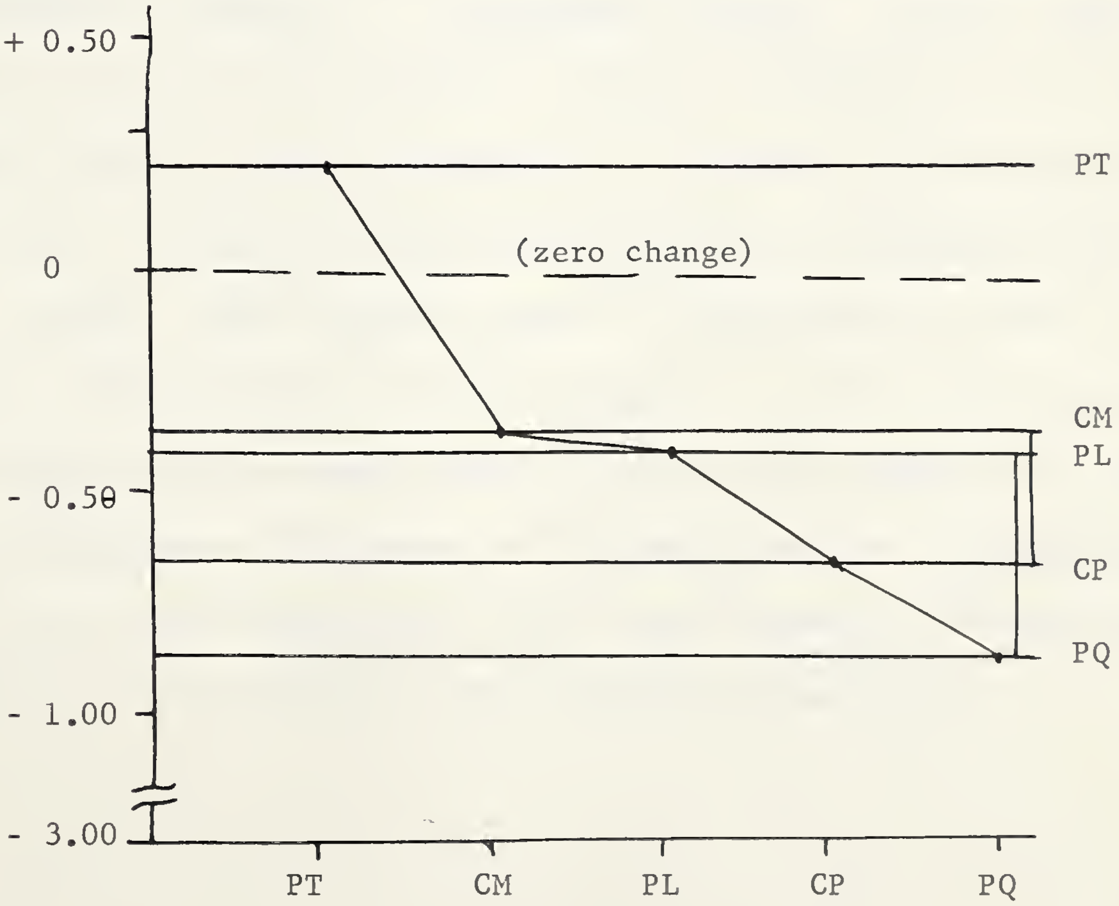
It should be admitted that, working at the .01 level, were an hypothesis to state that one mean is not significantly higher than another, one would, using this table, be drawing inferences at the .005 level of significance (i.e., a one-tailed test).

Table 40 indicates that there is a significant difference between D-score means for evidence of preparation for teaching and every other component category evaluated. In

addition to this, there is a significant difference between effectiveness of classroom management and effective use of personal qualities. That is, five of the ten differences tested earlier with analysis of variance, are so great as could occur under the null hypothesis in fewer than one trial out of 100. The findings shown in Table 40 are illustrated in Figure 6. In Figure 6 the means are plotted against a scale of D-scores and connected with a broken line. The distance between any two means may be seen by the distance between the horizontal lines extending from the ordinate through each mean. The two vertical bands represent the subsets within which there are no significant differences.

Figure 6

SIGNIFICANT AND NONSIGNIFICANT DISTANCES BETWEEN MEANS OF THE COMPONENTS OF TEACHING EFFECTIVENESS



It is interesting to compare the Duncan method with the conventional procedure of calculating a common standard error of the difference between any two means and multiplying that by a constant student's _t_ to obtain the minimum difference that is significant (see, e.g. 1, pp. 274 ff). The data were recalculated in the conventional manner, using the .01 level of significance. In this test, there were six significant differences rather than five. In addition to those already reported as significant, the difference between presentation of lessons (PL) and personal qualities (PQ) was found to be significant. In Figure 10 this is shown as the greatest non-significant difference. The conventional procedure would have excluded PL from the subset of non-significant mean differences PL-CP-PQ.

From the Duncan multiple range test of significant differences between k = 5 treatment means, one may infer that, within a group of 78 student teachers submitted to a failure-stress stimulus, one may expect to witness a difference in the extent to which several components of teaching effectiveness change. The subject tends to deteriorate in most components, particularly so with respect to the use of personal qualities and to a much lesser extent with respect to classroom management. Evidence of preparation for teaching may tend to be more successful than other aspects, actually improving slightly during the stressful experience.

REFERENCES FOR CHAPTER XVII

(1) Garrett, H. E. Statistics in Psychology and Education. (4th ed.). New York: Longmans Green, 1953.

(2) Edwards, A. L. Experimental Design in Psychological Research (Revised ed.). New York: Rinehart and Co., 1960

CHAPTER XVIII

SUMMARY AND CONCLUSIONS

Theory of Stress-Inducing Supervision Restated

The two-dimensional theory of interpersonal stress dynamics, developed and tested in this thesis, states that the optimum effectiveness of an individual working in an institutional environment is achieved through a unique combination of (1) the pressures, demands and expectations of the institution with (2) his own personality characteristics. These two elements, or dimensions, may be referred to as the external or objective drive stimuli and the internal or subjective drive arousibility factors. If each of these two dimensions could be quantified, one could probably show that the optimum proportions of each would vary from one individual to another. Lacking such quantification, particularly with respect to the external dimension, it was hypothesized that a combination of some quantifiable personality elements with a reasonably constant external stimulus situation would, when the external stimulus is intensified, result in quantifiable variations in behavior.

If optimum effectiveness depends upon a unique combination of the two dimensions, and if it can be supposed that several individuals are not operating as if that unique combination is obtained, then the hypothesis must predict differential results. Because of subjective factors, some individuals would be expected to react to an intensification

of the objective drive stimulus as though the interaction of the two resulted in that "unique combination" productive of optimum effectiveness. Their behavior would improve. Others, for the same reasons, would be expected to react as though the interaction of the two resulted in a combination that was quite in excess of the optimum. Their behavior would deteriorate. From *a priori* grounds, it should be expected that a marked increase in the objective drive stimulus through, for example, ego-threatening supervision, would produce at least some deterioration in effective reactions of the average individual.

This general statement must be surrounded by several qualifications and assumptions. The statement is intended to refer specifically to superior-subordinate relationships in an activity of individual supervision. The behavior to which it refers is classroom teaching behavior and this, it was assumed, is an activity that requires the teacher to use her personality resources in performing a complex task. It was assumed that individual supervision -- when the individual supervisor works directly with the individual teacher -- qualifies as part of the total objective drive stimulus for the teacher. When the supervisor acquaints the teacher with the institution's expectations for her general behavior relative to actual performance at the time, the event is one of evaluative interaction which increases the intensity of the objective drive stimulus for the teacher. The basis for this assumption rests chiefly with Selye's concept of emotional tension as a stress agent and with Rogers' concept of emotional

interactions arousing emotional tensions when an inconsistency with self-concept is experienced.

Although there may be a large number of personality elements entering the subjective drive arousibility factors, those that were incorporated into the theory -- neuroticism (including neurotic anxiety), general intelligence, and the achievement motive -- were selected for their pertinence to the objective aspects of the event with which they were expected to combine to form the total effective drive potential for the individual.

Probably no theory of behavior is ever completely new. The antecedents of the present theory can be traced to a number of sources including Getzels, Eysenck, Selye, and Rogers. Less directly, the influence of Sullivan, Parsons, Spence and others has been felt. In fact, the notion that behavior results from an interaction of the individual and his interpersonal environment is the private property of no single scientist or group. One theoretical setting of the present investigation is the theory of administrative relationships as a social system which was developed and tested by Getzels and his associates, notably Guba and Bidwell. The chief dimensions of the social system were the elements of the institution (roles and role-expectations) and of the individual (personality and need-dispositions). Observed behavior was a function of a combination of these two dimensions, specifically of the role-expectations and need-dispositions. An expansion of Getzels' theory was achieved in

the present thesis in four ways: (1) the personal dimension (Getzels called it idiographic) was re-examined and hypothesized to contain several personality elements which cannot adequately be subsumed under the term need-disposition, (2) the effect of stress resulting from expectations-personality incongruence was hypothesized to operate in more than one direction, (3) the degree and direction of change in behavior was hypothesized to be a function of the unique combination not only of the two major dimensions but also of a combination of elements within the personality dimension, and (4) the interpersonal dynamics of superior-subordinate stress were placed on a secure basis of theory and research. This expansion was accomplished chiefly by an attempt to incorporate into the Getzels theory the concepts and findings of the persons named earlier in this paragraph. The result was the theoretical framework presented in Chapter Eight which has been restated in condensed form in these pages.

Hypothetical Solutions

The basic notions of the theory as developed were re-cast into ten basic hypotheses which may be classified as follows:

1. General Effect (Hypotheses One, Two, and Three): When an acting faculty consultant, who bears essentially the same supervisory relationship as a regular faculty consultant to a student teacher, observes the latter teaching two consecutive lessons but, between the two, provides a strong

failure-stress stimulus by verbally comparing the subject's teaching to her personal disadvantage with institutional expectations, the effect upon classroom teaching performance during the second lesson will vary from one teacher to another. Because of individual differences in the personalities of the subjects, some will react as if the stimulus were a mild motivator and will improve their teaching noticeably although to a degree that will vary from one to another. Others, acting as if the stimulus together with their personal drive level produced an excessive total drive potential will deteriorate; some will deteriorate to a greater degree than others. With still other subjects, the combination of the stimulus and their personality characteristics results in too little change to be perceptible to trained observers. These changes -- positive, negative and zero -- in teaching effectiveness, when quantified for each subject, will be normally distributed throughout the total group, but the general or mean tendency of the group will be toward a decrement in general teaching effectiveness.

2. Effect of Personality Differences (Hypotheses Four, Five, Six, Seven, Eight and Nine): There are three personality variables -- neuroticism, general intelligence, and the achievement need -- each of which contributes independently to the general effect stated above. This may be stated as an inverse variation: the level of teaching effectiveness under stress, compared with control conditions, will vary inversely with the extent to which the individual possesses

each of the three personality variables. The lower the level of the variable, the higher will be the net effectiveness (in fact, an effectiveness increment from control conditions); the higher the level of the variable, the lower will be the net effectiveness (a decrement from control conditions). Furthermore, when the levels of personality variables are combined to form levels of drive arousibility intensity, the inverse variation is seen as a decreasing monotonic sequence from high net effectiveness for a low intensity level, to low net effectiveness for a high intensity level, the relationship being linear. When the relationship is stated in terms of the middle-range level of drive arousibility necessary for optimum effectiveness, the relation is seen as curvilinear: if level of total drive intensity is increased by the external stimulus, those who initially were below the middle-range move up toward the optimum under stress and, hence, improve; those at or above the middle-range move farther up under stress and, hence, farther from the optimum thereby deteriorating in performance.

3. Effect upon Components of Teaching (Hypothesis Ten): With respect to such components of teaching as use of personal qualities, evidence of preparation for teaching, presentation of lesson, contact with pupils, and classroom management, the effects of stress-inducing supervision will vary in the same manner as it does with respect to general teaching effectiveness.

General Results

The hypotheses were tested in a series of controlled experiments, one with each of 78 student teachers during their regular teaching activities in the public schools, who were visited by experimenters under the guise of acting faculty consultants who supervise this teaching activity. The criterion measure, or difference-rating (D-score), was the experimenter's assessment of the criterion variable: degree and direction of change in teaching effectiveness from the control to the experimental lesson. The hypotheses of general effect were tested with simple statistics of distributions. The hypotheses of effect of personality differences generally and upon the components of teaching were tested with sixteen analyses of variance. Eight of these were one-way designs: between means of subjects classified by levels of each of the predictor variables, between means of subjects classified by combined level of drive arousibility intensity, between linearity of trend and deviations from linearity, and between general means of components of teaching. Eight were two-way factorial designs: two-variable classification of D-scores for general effectiveness, a hypothetical replication of the same, an approximation of the same based on cell means, and five two-variable classifications for each of the components of teaching activity. No higher-dimension analyses were found necessary.

The results of these tests showed that the obtained research evidence supported the hypotheses of general effect,

supported the hypotheses of effects of personality differences with the exception of the need for achievement factor, and provided very limited support for the hypotheses of effect upon component aspects of teaching. The effect of stress upon components of teaching varied from one component to another but did not appear generally to be determined by differences in personality.

The investigator earlier decided on defensible *a priori* grounds to draw statistical inferences with ninety-five per cent confidence. In the majority of instances in which the null hypothesis was rejected, however, it was done so with a probability of error of less than one per cent. Similarly, when an acceptance of the null hypothesis was indicated, the probability of error was considerably below ninety-five per cent.

General Conclusions

The results lead to the conclusion that the theory of interpersonal stress in administrative relationships provides a valid means of predicting and interpreting a subordinate's behavior under conditions of administration-induced stress. The exceptions to this conclusion, so far as the present investigation was designed, are that the need for achievement failed as a predictor and that the remaining predictors did not differentiate between changes in effectiveness within the component aspects of teaching. These exceptions point to a weakness either in the theory or in the experimental

design. That the latter explanation is plausible with respect to the need for achievement was shown in Chapter Fourteen -- the subjects were homogeneously selected on this variable from the beginning. It is also possible that the method for arriving at sub-criterion measures of components of teaching was inadequate: the amount of observation time may have been sufficient for a trained observer to give a global assessment of teaching but not nearly enough for making more refined judgments on each of five component categories. These two exceptions notwithstanding, the investigation resulted in the administratively and psychologically valable conclusions that:

A. the method of supervision does have a significant effect upon teaching behavior,

  1. the nature of the effect is predictable when elements of the teacher's personality are known,

    (a) bright neurotics deteriorate significantly when the supervision is stress-inducing,

    (b) the less intelligent but more stable personalities improve under conditions that most would regard as stressful,

    (c) those in the middle range deteriorate only slightly;

B. this effect can be explained in terms of the teacher's drive level (combined personality elements) and the drive potential of the setting, including the direct supervision,

  1. under conditions of stress, net effectiveness shows an inverse linear relationship to level of drive arousibility intensity,

  2. the relationship of total effective drive potential (internal factors plus objective stimulus) is curvilinearly related to effective reaction potential: when total effective drive

potential is increased through stress-inducing supervision those that were initially low on this variable move upward toward to optimum level and those that were initially high move down from the optimum.

C. stress-inducing supervision affects several components of the teaching process differentially,

  1. the more personal aspects of teaching (use of personal qualities, contact with pupils) deteriorate the most,

  2. the more 'bookish' aspects of teaching (evidence of preparation for lesson, lesson plans, evidence of subject-matter knowledge) deteriorate the least and, in fact, improve,

  3. the more traditionally pedagogical aspects of teaching (classroom management, presentation of lesson) deteriorate to an extent somewhere between the other two, approximately to the same degree as general teaching effectiveness,

  4. the more intelligent the teacher, the more she deteriorates in effectiveness of classroom management and contact with pupils.

A Summary of Specific Findings

1. Administration of the Experiment.

   a. It was possible to conduct a controlled-experiment in a real-life teaching situation; security was maintained.

   b. The experiment achieved its objective of being stressful at least to a number of subjects who reported so verbally; three became obviously emotional, many more exhibited noticeable tension reactions.

2. General Effect on Change in Effectiveness.

   a. Criterion measures (D-scores for General Effectiveness) indicating degree and direction of change in teaching behavior from the control lesson to the experimental lesson did vary in both degree and direction.

   b. The distribution of D-scores did not significantly depart from a normal distribution having the same mean, range and variance.

c. The mean effect of the experiment was a significant decrement in teaching behavior, although the decrement was slight.

3. Effect of Personality Variables, Tested Singly.

   a. D-scores of subjects classified into three levels of neuroticism alone showed significant mean differences; the higher the neuroticism the greater the decrement, the lower the neuroticism the greater the increment.

   b. Mean D-scores of subjects classified into three levels of scholastic aptitude differed significantly; the higher the scholastic aptitude the greater the decrement, the lower the scholastic aptitude the greater the increment.

   c. Mean D-scores of subjects classified into three levels of need for achievement showed no significant difference.

   d. Mean D-scores of subjects classified into three levels of need for achievement via conformity showed no significant difference.

   e. Mean D-scores of subjects classified into three levels of need for achievement via independence showed no significant difference.

   f. The obtained F-ratio of less than unity for 3c, 3d, and 3e may be interpreted as indicating a sampling bias on this variable.

4. Effect of Personality Variables, Tested Independently.

   a. Mean D-scores of subjects classified into three levels of neuroticism with the effects of scholastic aptitude held constant showed significant differences in the same manner as 3a.

   b. Mean D-scores of subjects classified into three levels of scholastic aptitude with the effects of neuroticism held constant showed significant differences in the same manner as 3b.

   c. Interaction between the effects of neuroticism and scholastic aptitude is not significantly different from zero; the two vary independently; cell means were determined largely by marginal means.

d. All assumptions of homoscedasticity were met, in the tests so far.

e. The assumption of normality of distribution was justified in all but one of the 25 distributions (1 at 78; 15 at 26; 9 at approximately 9) so far, the lone exception being in a 3x3 lay-out where re-analysis showed its effects to be nonsignificant.

f. A 3x3 analysis of variance based on cell means accentuated the effects.

5. Trend of Combined Main Effects.

a. Categories of neuroticism and scholastic aptitude can be combined to form five levels of drive arousibility intensity to provide meaningful results.

b. There were significant differences among the means of subjects classified into five levels (high, medium high, medium, medium low and low) of drive arousibility intensity.

c. These differences were consistent, showing a trend from a high positive mean D-score for low category subjects to a low negative mean D-score for high category subjects.

d. The trend in 5c was linear, drive arousibility intensity being related to mean D-scores for general effectiveness in a decreasing monotonic sequence.

e. The farther was a drive arousibility category from the mid-point on either side, the greater was the mean D-score indicating that extremes on these personality variables improved or deteriorated to a greater degree than did those nearer the mean.

f. Thus, the relationship between total effective drive potential and effective reaction potential is curvilinear.

g. The significance of linearity findings were statistically very high.

6. Personality and the Components of Effective Teaching.

a. D-scores showing change in the effective use of personal qualities (PQ) were not significantly affected by differences among the subjects in terms of neuroticism or scholastic aptitude levels.

b. D-scores showing change in effectiveness for evidence of preparation for teaching (PT) were not significantly affected by personality differences among the subjects in MPI-N or A.C.E.).

c. D-scores showing change in effectiveness of lesson presentation (PL) were not significantly affected by personality differences among the subjects (in MPI-N or ACE).

d. D-scores showing change in effectiveness of contact with pupils (CP) were significantly affected by differences among the subjects in terms of scholastic aptitude; the higher the ACE level the greater the deterioration, the lower the ACE level the lesser the deterioration.

e. D-scores showing change in effectiveness of contact with pupils (CP) were not significantly affected by differences among the subjects in terms of their neurotic tendencies (MPI-N).

f. D-scores showing change in effectiveness of classroom management (CM) were significantly affected by scholastic aptitude differences among the subjects; the higher the ACE level the greater the deterioration, low ACE levels showed a slight mean improvement.

g. D-scores showing change in effectiveness of classroom management (CM) were not significantly affected by differences in neuroticism.

h. None of the interactions between neuroticism and scholastic aptitude in 6a to 6g significantly departed from zero.

7. Different General Effects Among Components.

a. There were significant differences among the D-score means (N=78) of each of the five component aspects of teaching; stress-inducing supervision affected the several aspects of teaching differentially.

b. Of the ten differences (among five means) between components of teaching, five were significant when tested with Duncan's new multiple range test.

c. Of the same ten differences, six were significant when tested with the conventional *t*-test; the multiple range test is the more powerful of the two.

d. Among means of D-scores for PL, CP and PQ there were no significant differences; among means of D-scores for CM, PL, and CP there were no significant differences. All other differences were significant.

8. Effect on Personal, Pedagogical and Academic Aspects.

   a. Preparation for teaching mean D-score was much higher than the others and was positive, showing a mean increment.

   b. Personal qualities and contact with pupils mean D-scores were much lower than the rest showing marked deterioration.

   c. Mean D-scores for classroom management (-.37) and lesson presentation (-.41) were approximately the same as the mean D-score for general effectiveness (-.38).

9. Anxiety and Extraversion.

   a. Anxiety, as measured by the Taylor MAS, differentiated the D-scores of subjects classified high, medium or low on this variable in the same way as neuroticism (MPI-N), but to a degree that was statistically not so significant.

   b. The correlation between MAS and MPI-N scores was .775 which is similar to reported correlations of .77.

   c. The personality variable known as extraversion-intraversion, when measured by the MPI-E, was not related to the change in effectiveness of subjects submitted to failure stress.

CHAPTER XIX

IMPLICATIONS

Earlier, the purpose of this thesis was stated as an attempt to discover, with respect to teachers who differed in personal attributes, what happens to their classroom teaching while acting within an administrative climate perceived by the teacher to contain psychological stress. It is a very real problem, both practically and theoretically. To the practicing administrator it has implications for all aspects of the job of supervision: evaluating and reporting, building effective teacher-supervisor relationships, maintaining a school climate conducive to learning, differentiating supervisory procedures according to individual teachers' needs and abilities, holding consultations, motivating teachers, and guiding projects of in-service education. To the social scientist it has implications for developing useful conceptualizations of interpersonal behavior, defining role-relationships and their limitations, strengthening existing concepts about administrative behavior and offering new ones, developing research to test useful hypotheses in real-life situations, and for demonstrating the need to call on the ideas and evidence of a variety of fields in the social sciences in order more fully to conceptualize a problem in administrative relationships.

A number of these implications are discussed in the present chapter. It must be cautioned that the reporting of

really secure findings ended with Chapter Eighteen and that the following sections refer only to what is implied by or inferred from the conclusions of the investigation.

Implications for Improved Educational Practice

A recent statement that draws attention to the two chief roles that a supervisor must accept if he is to improve instruction recognizes that teachers improve in two ways:

> . . . there are two major ways by which teachers improve. One way is by learning more about teaching and the subjects taught. The other is by being challenged by the job and being stimulated to put into practice their present knowledge of good teaching. (1, p. 10)

According to this statement, the supervisor is both a consultant and a motivator. The present study contains several implications for the effective performance of these supervisory functions. Common to them all is the inference that the supervisor must differentiate the performance of his role to suit the individual needs and abilities of those in his charge. In consultation, for example, the non-neurotic will tend to see the "job" as the point of reference whereas the neurotic tends to take the "self" or "my work" as the point of reference. The same consultative method by a supervisor may, therefore, have different results with two different teachers.

In a sense, any supervisory activity that tends to increase the drive level of the teacher performs the motivational

function of supervision. Drive, however, is non-directed whereas motivation is customarily thought of as directed. The problem of supervision, it may now be seen, is one of providing a stimulus that will raise low-drive personnel to the optimum level without raising them or others past it. It was seen that negative supervision raised some persons to the optimum (or so it was inferred), and that it also raised persons with a higher drive intensity level past the optimum to a point of deteriorated performance. Effective supervision will use a different form of stimulation, one that -- according to Carl Rogers -- involves a complete absence of ego-threat. One way of doing this is to role-educate the teacher or the staff, through the consultative function, to the point where a perceived conflict (non-ego threatening) exists between the new concept of the job that is now accepted and the old concept of the job as it had been performed. One would predict that if the challenge of the job appears greater and if no threat is involved, the resulting tension or stress (in mild degree, the terms are accurately used) will energize the teacher to improve performance. The direction of energy is provided by the nature of the stimulus -- the broader and more demanding concept of the teaching task. This is not a novel inference; it has been the pattern of effective in-service education projects for years. The supervisor has evaluated the performance of his staff in general terms, has sought the cooperation of outstanding staff members in planning a project such as a "resource" unit, has encouraged groups

of teachers to work in committees on the project in which they -- by their own efforts -- learn much more about what they should be teaching and how they could be teaching it. Ultimately, the new learning becomes put into practice, the motivation has been increased and few, if any, participated under threat. A less effectively operated facsimili of the same project is one that requires teacher-participation, and where supervisory personnel acquaint the staff of their role-expectations directly through addresses or pseudo-discussions. The element of threat cannot be ignored, neither can the implied evaluation of general performance; what the effects of such a program upon the teaching activity of intelligent and sensitive teachers would be must remain a conjecture. The present theory and investigation, however, would suggest a decrement.

To carry the inference a step further is to speculate on the effects upon instruction of a school principal who, perhaps unwittingly, maintains his staff in a state of continual tension. Not only would the present theory suggest that effectiveness of instruction in that school will be less than optimum, but also it suggests that it is the personal or "warm" aspects of teaching that will suffer in that school. The principal may, of course, take some solace in the observation that "high standards" are being maintained because of the tendency of teachers under tension to rely more heavily upon the more bookish aspects of teaching: evidence of lesson preparation and rigid planning which are more obvious to the

casual visitor to a classroom. In short, the kind of "social climate" or "interpersonal climate" of the staff that the principal effects can be a major result of his general supervisory procedure.

Because any situation of individual supervision may be stressful to a teacher and because the extent and effects of this stress will vary greatly from one teacher to another depending upon their respective levels of drive arousibility intensity, the validity of formal inspections may be called in question. A formal inspection has as its major purpose the evaluation of teaching effectiveness. By itself it is not a supervisory procedure but an administrative device. Its purpose is to acquaint the administration with the standard of performance within the system generally and for each teacher. Not until remedial or motivational action is taken is the practice one of supervision. Because of personality differences among teachers, however, the inspector (the term here refers to function, not to position) may not be evaluating a typical sample of a teacher's behavior. Some teachers may teach more effectively, some much less effectively, during formal inspection than during normal classroom activity. In fact, their behavior may be presumed to follow closely the results of this investigation. An evaluation is generally more valid when a larger sampling of behavior is taken, that is when it is based on longer visits or more frequent visits in the instance of formal inspection. But this principle of evaluation does not necessarily hold for the present example if neither the teacher nor the

inspector have altered their respective personalities during the interim. The validity of a rating is generally enhanced when more than one person does the rating, that is when there is more than one 'inspector'. It has been shown (3), however, that evaluators themselves differ widely in their respective evaluations of the same performance. Indeed, the findings of the present study imply that if two separate evaluators were perceived by a teacher as representing different degrees of ego-threat there would be a difference in the total effective drive potential of the situation for that teacher and that this, in turn, would result in differential effec ts upon her classroom teaching behavior. The implication is that formal ratings are of doubtful validity.

Perhaps the most important practical implication is the need for what has been called "differential supervision" (2). Just as the weight of research on the teaching-learning process has led teachers to observe individual differences among pupils and adapt their instruction accordingly, so now the research on specific administrative relationships indicates the need for supervisors to be alert to individual differences in personality and aptitude among their teachers and to individualize their supervisory procedures. The differentiated approach implies that a supervisor must be skilful in a variety of procedures. The appropriate application of these procedures, however, is somewhat more difficult for the supervisor than for the teacher. The classroom teacher providing differential instruction is guided by a large battery of test and record

information on each of the pupils in her charge. Customarily the supervisory person has little or no information of this nature regarding his staff; this is particularly so with respect to the predictor variables used in the present investigation. As one consults with a staff member, the answers to questions of this person's anxiety, neuroticism, intelligence, or general drive arousibility can only be inferred, and such inferences made by untrained persons may frequently be in error. Here, then, lies an implication for the training and selection of persons for positions of administrative responsibility: that bases for selection of such persons include the successful conduct of interpersonal relations and that programs of administration education include the alerting of candidates to the personal factors among the supervised that cause them to react differentially to supervisory procedures. In all likelihood, these factors are more numerous than those that have been isolated and studied in the present investigation.

Theoretical Implications

A reasonable inference is that what was found to be true of student teachers under the experimental conditions would be true of teachers under similar conditions. Although the entire superintendent-teacher relationship was not replicated, the essential elements of superior-subordinate relationships were established, in a process of individual supervision. Regularly employed teachers, of course, may react not so immediately or acutely because of their greater maturity

and familiarity with routine.

It is also reasonable to infer that the same results would be true of a learning situation in so far as the subjects of this experiment were engaged in a complex task. Although the experimenter-subject relationship was not quite the same as a teacher-pupil relationship, there were present sufficient essential elements both in the relationship and in the task being performed for the investigation to be generalized to problems of learning.

Although it was stated more than once that several personality variables probably enter a situation to produce the individual's drive arousibility intesnity, the investigation has made it possible to infer that there are at least two very important determiners of effective reactions under conditions that many would regard as stressful. These two, in terms of their tests, are Neuroticism and Scholastic Aptitude or General Intelligence.

Thus it is possible to see implications of the present findings for three large fields of interest: administrative relationships, learning theories, and personality dynamics. Briefly, the implications are these: (1) Getzels' theory of administrative relationships is a promising theory which contains either several weaknesses of fact or ambiguities of expression. The only way of resolving the present findings with Getzels' theory of administration relationships is to claim that our subjects who improved their teaching somehow

had a "need to be stimulated in this manner" so that the treatment fulfilled the need. There are two drawbacks to this: (a) it fails to explain why persons with low general intelligence and less than average neuroticism have a "need-disposition" requiring failure-stress, and (b) it makes it possible to justify any need-theory by inventing after-the-fact needs for whatever behavior was observed. In the present instance, experiment followed theory-development and the specification of variables. Nevertheless, the fact that Getzels specified some of the particular dimensions of institution-personality conflict, and that the general relationship was useful for the present study, suggestes that the theory deserved further attention. (2) Eysenck's hypothesis that an individual's level of neuroticism enters into his total drive level to help determine his success or failure in complex task performance was decidedly borne out in the present investigation. This implies that the value of Eysenck's theories of personality and learning merits further consideration. It is interesting, however, that Eysenck did not himself include "general intelligence" in his hypothesis (which was an adaptation of the Yerkes-Dodson Law); his own personality theory posits the cognitive aspect of personality as a very important one, and one which constantly interacts with the conative aspects of neuroticism and extraversion-introversion. (3) The chief implications for personality theories are that (a) scholastic aptitude or general intelligence is an important determinant of behavior either because of its specific cognitive denotations or because of its affective and conative

connotations, and (b) the self-concept theory of personality, when expanded to include self-expectations relative to the achievement of success, can provide fruitful insights into problems of behavior under psychological stress. Both of these implications share the concept, employed by Getzels, that effective behavior is a function of the expectations and pressures of the external interpersonal world interacting with the individual's unique personality. The present investigation, however, makes it possible to add the words "interacting with the relevant aspects of the individual personality." By this it is inferred that a combination of those personality aspects that are relevant to the situation will determine adjustment to or success in the situation. Extraversion, for example, was probably not so relevant to the stress situation as was neuroticism. Elsewhere (2), it has been shown that neuroticism is not so relevant to regular student teaching as is general intelligence. It would appear that the leader of a group must differentiate his leadership styles and procedures so that the external factors and the individuals' characteristics may be combined to produce optimum effectiveness. That, in a very general manner, is what is implied by the term "differential supervision."

Implications for Future Research

The present investigation was designed more as an adventure into new ways of conceptualizing old problems, than as a means of solving them. A considerable amount of research therefore, will be required before definitive answers, if ever,

can be given. The two avenues that this research would take are to expand and to intensity.

The need for expansive research is implied by the specific nature of the problem investigated in this thesis: one kind of supervision, a very few personality variables and immediate effects. Several different kinds of supervisory situations, both with and without ego-threat, require investigation under hypotheses that specify the administratively or psychologically significant aspects of the organizational environment. Personality variables must include those at the periphery of personality structure -- specific habits, traits, interests -- as well as the more basic dimensions in order to observe more clearly the probable effects upon certain aspects of performance. Whereas immediate effects, only, were investigated in the present thesis there is a need to test hypotheses predicting long-term effects upon performance. There is little to be learned from the present study if it can be shown that two hours after the supervisor's visit the effects have "worn off" and left no trace. On the other hand, the effect may well be intensified upon each successive visit or, as in the instance of a principal with his staff, with each daily reminder that the principal is a stress-figure. Long-term effects under a variety of circumstances require careful investigation. The value of the present study was that it suggested questions that must be answered, and a few ways of phrasing them.

The need for intensive research is implied by the nature of the variables employed in this study: one may refer to them loosely as drive, intelligence, teacher effectiveness, neuroticism, anxiety, stress and so forth. If anything, the present investigation has served to raise some questions pertaining to the nature and effects of these variables. It would appear that some aspects of effective teaching are affected differently from others; then what is effective teaching in terms of its components? in terms of its results? Questions about the non-unitary nature of measured intelligence and measured anxiety have been raised. The introduction of variables that were not included in this study, the clearer specification of experimental conditions and of criterion measured would permit -- with a sufficiently large sample -- a major investigation to be carried out using the methods of factor analysis. Factor analysis, which is a convenient method of classifying the behavior of persons under test situations, would make it possible to sort those variables (or aspects of those that are not clearly defined) that are significant determinants of behavior under specified supervisory situations from those that are not.

Short of the preceding more ambitious research implications there is, of course, a variety of questions requiring clarification. What is the nature of supervisor-supervisee communication? Does the latter perceive as ego-threatening an intended supportive statement from the former? Is the

teaching load, in certain systems, already sufficiently heavy that projects of in-service education simply create stress? Is there a "teacher personality" from which supervisors may generalize their procedures? Do veteran teachers differ in reactions to stressful supervision from novices? With which reference group -- parents, pupils, colleagues, superiors -- do teachers encounter the most stress-inducing situations? Further research on these and similar problems will be needed the more clearly to understand the dynamic nature of administrative relationships and its effects upon attaining the goals of education.

REFERENCES FOR CHAPTER XIX

(1) Andrews, J. H. M. and W. H. Worth "The Principal as Consultant and Motivator," The ATA Magazine 41:10 (June 1961) pp 10-13.

(2) Brown Alan F. "Differential Supervision I. Personality and Effective Teaching." The C.S.A. Bulletin, 1:1 (Sept. 1961) pp. 60-70.

(3) Worth, W. H. "Can Principals Rate Teachers?" The C.S.A. Bulletin, 1:3 (Jan. 1962) in press.

BIBLIOGRAPHY

Ahrend, M. R. "Helping Teachers with Their Tensions," Childhood Education, 28 (Jan. 1952) 211-215.

Aikenhead, J. D. "Teacher Satisfactions and Discouragements," Alberta Journal of Educational Research, 4:2 (June 1960) 92-102.

American Council on Education. Manual of Instructions for the Psychological Examination for College Freshmen. Princeton, N.J.: Cooperative Test Division, Educational Testing Service, 1950.

"Anatomy of Angst, The," Time Magazine, 77:14 (Mar. 31, 1961) 46-51.

Andrews, J. H. M. "Administrative Significance of Psychological Differences between Secondary Teachers of Different Subject Matter Fields," Alberta Journal of Education Research, 3:4 (Dec. 1957) 199-208.

----------"A Deterrent to Harmony Among Teachers," Administrator's Notebook, 4:7 (Mar. 1958) 1-4.

----------and Alan F. Brown "Can Principals Exclude Their Own Personality Characteristics When They Rate Their Teachers?" Educational Administration and Supervision, 45:4 (July 1959) 234-242.

----------and W. H. Worth "The Principal as Consultant and Motivator," The A.T.A. Magazine, 41:10 (June 1961) 10-13.

Ausubel, D. P., H. M. Schiff and M. Goldman "Quantatitive Aspects in the Learning Process Associated with Anxiety," Journal of Abnormal and Social Psychology, 48 (1953) 537-547.

Axelrod, H. S., G. L. Cowen and F. Heilizer "The Correlates of Manifest Anxiety in Stylus Maze Learning," Journal of Experimental Psychology, 51 (1956) 131-138.

Barnard, Chester. The Functions of the Executive. Cambridge: Harvard University Press, 1938.

Barr, A. S., D. E. Eustice and E. J. Noe "The Measurement and Prediction of Teacher Efficiency," Review of Educational Research, 25:3 (June 1955) 261-269.

Beam, J. C. "Serial Learning and Conditioning under Real Life Stress," Journal of Abnormal and Social Psychology, 51 (1955) 543-551.

Bendig, A. W. "Age, Sex, and the Manifest Anxiety Test," *Journal of Consulting Psychology,* 18 (Feb. 1954) 16.

----------"Extraversion, Neuroticism and Manifest Anxiety," *Journal of Consulting Psychology*, 21:5 (Oct. 1957) 398.

Bevan, W. and R. A. Maier "Emotional Tension and the Generality of its Effect upon Intellectual Performance," *Journal of Personality*, 26:3 (Sept. 1958) 330-336.

Bidwell, C. E. "The Administrative Role and Satisfaction in Teaching," *Journal of Educational Sociology*, 39 (Sept. 1955) 41-47.

---------- "Some Causes of Conflict and Tension Among Teachers," *Administrator's Notebook*, 4:7 (Mar. 1956) 1-4.

----------"Administration and Teacher Satisfaction," *Phi Delta Kappan*, 37 (Apr. 1956) 285-288.

Brayfield, A. H. and W. H. Crockett "Employee Attitudes and Employee Performance," *Psychological Bulletin*, 52:5 (1955) 397-424.

Britterman, M. E. and W. H. Holtzman "Conditioning and Extension of the Galvanic Skin Response as a Function of Anxiety," *Journal of Abnormal and Social Psychology*, 47 (1952) 615-623.

Broadhurst, P. L. "Determinates of Emotionality in the Rat, I: Situational Factors," *British Journal of Psychology*, 47 (1957) 1-12.

----------"Determinates of Emotionality in the Rat, II: Antecedent Factors," *British Journal of Psychology*, 49 (1958) 12-20.

Brookover, W. B. "Research on Teacher and Administrator Roles," *Journal of Educational Sociology*, 29:1 (Sept. 1955) 2-13.

Broverman, D. M. and R. S. Lazarus "Individual Differences in Task Performance under Conditions of Cognitive Interference," *Journal of Personality*, 26:1 (Mar. 1958) 94-105.

Brown, Alan F. "The Self in Interpersonal Theory: The Relation between Perceptions of Self and of Significant Other," *Alberta Journal of Educational Research*, 3:3 (Sept. 1957) 138-148.

----------"Differential Supervision: I. Personality and Effective Teaching," *The C.S.A. Bulletin,* 1:1 (Sept. 1961) 31-35.

----------and J. H. M. Andrews "Can Principals Exclude Their Own Personality Characteristics when They Rate Their Teachers?" *Educational Administration and Supervision*, 45:4 (July 1959) 234-242.

Buchwald, A. E. and H. G. Yamaguchi "The Effect of Change in Drive Level on Habit Reversal," *Journal of Experimental Psychology*, 50 (1955) 265-268.

Buss, A. H., M. Wiener, A. Durkee and M. Baer "The Measurement of Anxiety in Clinical Situations," *Journal of Consulting Psychology*, 19:2 (Apr. 1955) 125-129.

Campbell, M. V. "Teacher-Principal Agreement on the Teacher Role," *Administrator's Notebook*, 7:6 (Feb. 1959) 1-4.

Campbell, R. F. and J. M.Lipham (Eds.) *Administrative Theory as a Guide to Action*. Chicago: Midwest Administration Center, 1960.

Castaneda, A. and D. S. Palermo "Psychomotor Performance as a Function of Amount of Training and Stress," *Journal of Experimental Psychology*, 50 (1955) 175-179.

Charles, H. and E. Mech "Performance in an Operational 'Stress' Situation Related to a Projective Technique," *Journal of Educational Research*, 48 (Mar. 1955) 509-519

Chase, F. S. "Factors for Satisfaction in Teaching," *Phi Delta Kappan*, 33:3 (Nov. 1951) 127-132).

----------"How to Meet Teachers' Expectations of Leadership," *Administrator's Notebook*, 1:9 (Apr. 1953) 1-4.

Clarke, S. C. T. "School Control and Discipline," *Leadership Course for Alberta School Principals*. Edmonton: Division of Educational Administration, University of Alberta, 1956.

Congreve, W. J. "Administrative Behavior and Staff Relations," *Administrator's Notebook*, 6:2 (Oct. 1957) 1-4.

Cowen, E. L. "The Influence of Varying Degrees of Psychological Stress on Problem-Solving Rigidity," *Journal of Abnormal and Social Psychology*, 47 (1952) 512-519.

----------, F. Heilizer, H. L. Azelrod and S. Alexander "The Correlates of Manifest Anxiety in Perceptual Reactivity, Rigidity and Self-concept," *Journal of Consulting Psychology*, 21:5 (Oct. 1957) 405-411.

----------, M. Weiner and J. Hess "Generalization of Problem-Solving Rigidity," Journal of Consulting Psychology, 17:2 (1953) 100-103.

Cronbach, L. J. Educational Psychology, New York: Harcourt Brace, 1954.

Deese, J. R. S. Lazarus, and J. Keenan "Anxiety, Anxiety Reduction and Stress in Learning," Journal of Experimental Psychology, 46 (July 1953) 55-60.

Diethelm, O. and M. R. Jones "Influences of Anxiety on Attention, Learning, Retention and Thinking," Archives of Neurology and Psychiatry. 58 (1947) 325-336.

Dollard, J. and N. E. Miller. Personality and Psychotherapy. New York: McGraw-Hill, 1950.

Edwards, A. L. Personal Preference Schedule and Manual (Revised Ed.). New York: Psychological Corp., 1959.

----------The Social Desirability Variable in Personality Assessment and Research. New York: The Dryden Press, 1957.

----------Experimental Design in Psychological Research (Revised Ed.). New York: Rinehart and Co., 1960.

Elizur, A. "A Content Analysis of the Rorschach with Regard to Anxiety and Hostility," Journal of Projective Techniques, 13 (1949) 247-283.

Eysenck, H. J. Dimensions of Personality. London: Routledge and Kegan Paul, 1947.

----------The Scientific Study of Personality. London: Routledge and Kegan Paul, 1952.

----------The Structure of Human Personality. London: Methuen, 1953.

----------A Dynamic Theory of Anxiety and Hysteria," Journal of Mental Science, 101 (Jan. 1955) 28-51.

----------"Reminiscence, Drive and Personality Theory," Journal of Abnormal and Social Psychology, 53 (1956) 328-333.

----------The Dynamics of Anxiety and Hysteria. London: Routledge and Kegan Paul, 1957.

----------"A Short Questionnaire for the Measurement of Two Dimensions of Personality," Journal of Applied Psychology, 42 (1958) 14-17.

----------The Maudsley Personality Inventory and Manual. London: University of London Press, 1959.

Farber, I. E. "Response Fixation Under Anxiety and Non-anxiety Conditions," Journal of Experimental Psychology, 38 (1948) 111-131.

----------and K. W. Spence "Complex Learning and Conditioning as a Function of Anxiety," Journal of Experimental Psychology, 45 (1953) 120-125.

----------"Anxiety as a Drive State," Nebraska Symposium on Motivation, (M.R. Jones, Ed.). Nebraska: University of Nebraska Press, 1954, pp. 1-45.

Ferguson, G. A. Statistical Analysis in Psychology and Education. New York: McGraw-Hill, 1959.

Ferneau, E. "Role-expectations in Consultations," Unpublished Ph.D. thesis, University of Chicago, 1954, reported in Hencley, S. P. "Leader Behavior in Social Systems," Alberta Journal of Educational Research, 7:2 (June 1961) 109.

Flanders, N. A. "Personal-social Anxiety as a Factor in Experimental Learning Situations," Journal of Educational Research 45 (Oct. 1951) 100-110.

Forlano, G. and H. C. Axelrod "The Effects of Repeated Praise or Blame on the Performance of Introverts and Extroverts," Journal of Educational Psychology 28 (1937) 92-100.

Franks, C. M. "Personality Factors and Rate of Conditioning," British Journal of Psychology. 48 (1957) 119-126.

Freeman, G. L. "Suggestions for a Standardized 'Stress' Test," Journal of General Psychology 32:1 (Jan. 1945) 3-11.

Fromm, Erich. The Sane Society. New York: Rinehart, 1955.

Gaier, E. L. "Selected Personality Variables and the Learning Process," Psychological Monographs, 66:17 (1952) Whole No. 349, pp. 28.

Galbraith, J. K. The Affluent Society. London: Hamish, 1958.

Garrett, H. E. Statistics in Psychology and Education,(4th ed.) New York: Longmans Green 1953.

Getzels, J. W. "A Psycho-sociological Framework for the Study of Administration," Harvard Educational Review, 22 (Fall 1952) 235-246.

----------"Administration as a Social Process," in Administrative Theory in Education (A.W. Halpin, Ed.). Chicago: Midwest Administration Center, 1958.

----------"Theory and Practice in Educational Administration: An Old Question Revisited," Administrative Theory as a Guide to Action, (R. F. Campbell and J. M. Lipham, Ed.). Chicago: Midwest Administration Center, 1960, pp. 37-58.

----------and E. G. Guba "Role, Role Conflict and Effectiveness: An Empirical Study," American Sociological Review, 19 (1954) 164-175.

----------and E. G. Guba "The structure of Roles and Role Conflict in the Teaching Situation," Journal of Educational Sociology, 29 (Sept. 1955) 30-40.

----------and E. G. Guba "Social Behavior and the Administrative Process," School Review, 64:5 (Winter 1957) 423-441.

----------and H. A. Thelen "The classroom as a Unique Social System," The Dynamics of Instructional Groups: The 59th N.S.S.E. Yearbook, Part II. Chicago: The Society: 1960, pp. 53-82.

Glixman, A. G. "Recall of Completed and Incompleted Activities Under Varying Degrees of Stress," Journal of Experimental Psychology 39:3 (June 1949) 281-295.

Goodstein, L. D. "Interrelations Among Several Measures of Anxiety and Hostility," Journal of Consulting Psychology, 18 (Feb. 1954) 35-39.

Gordon, W. M. and D. E. Berlyne "Drive-level and Flexibility in Paired Associate Nonsense-syllable Learning," Quarterly Journal of Experimental Psychology, 6 (1954) 181-185.

Gorlow, L. C. N. Zimet and H. J. Fine "The Valdity of Anxiety and Hostility Rorschach Content Scores Among Adolescents," Journal of Consulting Psychology, 16 (Feb. 1952) 13-75.

Gough, H. G. California Psychological Inventory. Palo Alto: Consulting Psychologists Press, 1956.

Griffiths, D. E. _Administrative Theory_. New York: Appleton-Century-Crofts, 1959.

Guba, E. G., and C. E. Bidwell. _Administrative Relationships_. Chicago: Midwest Administration Center, 1957.

Guilford, J. P. _Fundamental Statistics in Psychology and Education_ New York: McGraw-Hill, 1956.

Hall, C. S. "Some Conditions of Anxiety Extinction," _Journal of Abnormal and Social Psychology_, 51 (1955) 126-132.

----------and G. Lindzey. _Theories of Personality_. New York: Wiley, 1957.

Halpin, A. W. (ed.). _Administrative Theory in Education._ Chicago: Midwest Administration Center, 1958.

Haugen, G. B., H. H. Dixon and H. A. Dickel. _A Therapy for Anxiety Tension Reactions_. New York: MacMillan, 1958.

Hebb, D. O. _A Textbook of Psychology_. Philadelphia: W. B. Saunders Co., 1959.

Heilitzer, E., H. S. Axelrod and E. L. Cowen "The Correlates of Manifest Anxiety in Paired Associates Learning," _Journal of Personality_, 24 (1956) 463-474.

Hilgard, E. R., L. B. Jones and S. J. Kaplan "Conditioned Discrimination as Related to Anxiety,"_Journal of Experimental Psychology_, 42 (1951) 94-99.

Hills, R. J. "A New Concept of Staff Relations," _Administrator's Notebook_, 8:7 (Mar. 1960) 1-4.

Hughes, J. G., J. L. Sprague and A. W. Bendig "Anxiety Level, Response Alteration, and Performance in Serial Learning," _Journal of Psychology_, 38 (1954) 421-426.

Janis, I. L. _Psychological Stress_. New York: Wiley and Sons, 1958.

Jersild, A. T. _When Teachers Face Themselves_. New York: Bureau of Publications, Teachers College, Columbia University, 1955.

Kendall, E. "The validity of Taylor's Manifest Anxiety Scale," _Journal of Consulting Psychology_ 18 (1954) 429-432.

Lazarus, R. S., J. Deese and S. Osler. "The Effects of Psychological Stress upon Performance." Psychological Bulletin. 49 (1952) 293-317.

----------, J. Deese, and R. Hamilton "Anxiety and Stress in Learning: The Role of Intraserial Duplication," Journal of Experimental Psychology, 47 (1954) 111-114.

----------and C. W. Eriksen "Effects of Failure Stress Upon Skilled Performance," Journal of Experimental Psychology, 42 (Feb. 1952) 100-105.

Lindquist, E. F. Design and Analysis of Experiments in Psych- and Education. Boston: Houghton-Mifflin, 1953.

Lindzey, G. and A. S. Newburg "Thematic Apperpection Test: A Tentative Appraisal of Some 'Signs' of Anxiety," Journal of Consulting Psychology, 18 (Dec. 1954) 389-395.

Lipham, J. M. "Personal Variables of Effective Administrators," Administrator's Notebook, 9:1 (Sept. 1960) 1-4.

Lotsof, E. J. and W. L. Downing "Two Measures of Anxiety," Journal of Consulting Psychology, 20 (June 1956) 170.

Lucus, J. "The Interaction Effects of Anxiety, Failure, and Intraserial Duplication," American Journal of Psychology, 65 (1952) 166-173.

MacArthur R. S. "The Coloured Progressive Matrices as a Measure of General Intellectual Ability for Edmonton Grade III Boys," Alberta Journal of Educational Research, 6:2 (June 1960) 67-75.

McClelland, D. C. Personality. New York: The Dryden Press, 1953.

----------Studies in Motivation. New York: Appleton-Century-Crofts, 1955.

----------and F. S. Apicella. A Functional Classification of Verbal Reactions to Experimentally Induced Failure," Journal of Abnormal and Social Psychology, 40 (1945) 376-390.

----------, J. W. Atkinson, R. A. Clark and E. L. Lowell. The Achievement Motive. New York: Appleton-Century-Crofts, 1953.

McCreary, J. B. and A. W. Bendig "Comparison of Two Forms of the Manifest Anxiety Scale," Journal of Consulting Psychology, 18 (June 1954) 206.

McKinney, F. "Emotional Factors in Learning," Journal of General Psychology, 9 (1933) 101-116.

McNemar, Q. Psychological Statistics (2nd ed.). New York: Wiley & Sons, 1955.

Maier, N. F. Frustration. New York: McGraw-Hill, 1949.

Mendler, G. and S. B. Sarason "A Study of Anxiety and Learning," Journal of Abnormal and Social Psychology, 47 (1952) 166-173.

Matarazzo, J. D. "MMPI Validity Scores as a Function of Increasing Levels of Anxiety," Journal of Consulting Psychology, 19:3 (June 1955) 213-217.

----------and J. S. Phillips "Digit Symbol Performance as a Function of Increasing Levels of Anxiety," Journal of Consulting Psychology, 19:2 (April 1955) 131-134.

May, Rollo. The Meaning of Anxiety. New York: Ronald, 1950.

Moldawsky, S. and P. C. Moldawsky "Digit Span as an Anxiety Indication," Journal of Consulting Psychology, 16 (April 1952) 115-118.

Montague, E. K. "The Role of Anxiety in Serial Rote Learning," Journal of Experimental Psychology, 45 (1953) 91-96.

Moser, R. P. "Leadership Patterns of School Superintendents and School Principals,: Administrator's Notebook, 6:1 (Sept. 1957) 1-4.

Mowrer, O. H. Learning Theory and Behavior. New York: Wiley, 1960.

----------"A Stimulus-Response Analysis of Anxiety and its Role as a Reinforcing Agent," Psychological Review, 46 (1939) 553-565.

Moyer, D. C. "Leadership that Teachers Want," Administrator's Notebook, 3:7 (Mar. 1955) 1-4.

Murphy, G. Personality. New York: Harper and Bros., 1947.

Murray, H. A. Explorations in Personality. New York: Oxford University Press, 1938.

O'Connor, J. P. and J. W. Stafford "Some Patterns of Manifest Anxiety," Journal of Clinical Psychology, 12:2 (April 1956) 160-163.

Osler, S. F. "Intellectual Performance as a Function of Two Types of Psychological Stress," Journal of Experimental Psychology, 47 (1954) 115-121.

Otto, H. J. and E. Ol Melby "An Attempt to Evaluate the Threat of Failure as a Factor in Achievement," Elementary School Journal, 35 (1935) 588-596.

Parsons, T. Essays in Sociological Theory Pure and Applied. Glencoe, Ill.: The Free Press, 1949.

----------and E. Shils. Toward a General Theory of Action. Cambridge: Harvard University Press, 1951.

Prince, R. "Individual Values and Administrative Effectiveness," Administrator's Notebook, 6:4 (Dec. 1957) 1-4.

Ramond, C. K. "Anxiety and Tasks as Determiners of Verbal Performance," Journal of Experimental Psychology, 46 (1953) 120-124.

Raphelson, A. C. "The Relationships Among Imaginative, Direct Verbal, and Physiological Measures of Anxiety in an Achievement Situation," Journal of Abnormal and Social Psychology, 54 (1957) 13-18.

----------and R. W. Moulton "The Relationship Between Imaginative and Direct Verbal Measures of Test Anxiety Under Two Conditions of Uncertainty," Journal of Personality, 26:4 (Dec. 1958) 556-567.

Robinson, H. A. "Job Satisfaction Researches of 1956," Personnel and Guidance Journal 36:1 (Sept. 1957) 34-37.

Rogers, C. R. Client-Centred Therapy. Boston: Houghton-Mifflin, 1951.

----------and R. F. Dymond Psychotherapy and Personality Change. Chicago: University of Chicago Press, 1954.

Roso, B. M., J. W. Rupul and D. A. Grant "Effects of Personal Impersonal and Physical Stress upon Cognitive Behavior in a Card-sorting Problem," Journal of Abnormal and Social Psychology, 47 (1952) 546-551.

Sarason, S. B. "Effects of Anxiety, Motivational Instruction, and Failure on Serial Learning," Journal of Experimental Psychology, 51 (1956) 253-260.

----------, G. Mandler and P. C. Craighill "The Effect of Differential Instructions on Anxiety and Learning," Journal of Abnormal and Social Psychology, 47 (1952) 561-566.

Sartre, J. P. L'Etre et le Neant (H. E. Barnes, trans.: Being and Nothingness). New York: Philosophical Library, 1956.

Schwab , R, and J. Pritchard "Situational Stresses and Extra-pyramidal Disease in Different Personalities." Life Stress and Bodily Disease, Assoc. Research on Nervous Mental Disease, 1949, pp. 48-60.

Selye, H. The Story of the Adaptation Syndrome. Montreal: Acta Inc., 1952.

----------The Stress of Life. Toronto: McGraw-Hill Book Co., 1956.

Selznick, P. "Foundations of the theory of organization," American Sociological Review, 13 (Feb. 1948) p. 29.

Sharma, C. L. "Who Should Make What Decisions?" Administrator's Notebook, 3:8 (April 1955) 1-4.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Co., 1956.

Silverman, R. E. "Anxiety and the Mode of Response." Journal of Abnormal and Social Psychology, 49 (1954) 538-542.

----------"The Edwards Personal Preference Schedule and Social Desirability," Journal of Consulting Psychology, 21:5 (Oct. 1957) 402-404.

Solomon, R. L. and L. L. Wynne "Traumatic Avoidance Learning: The Principles of Anxiety Conservation," Psychological Review, 61 (1954) 353-385.

Spence, K. W. "Learning and Performance in Eyelid Conditioning as a Function of Intensity of the UCS," Journal of Experimental Psychology, 45 (1953) 57-63.

----------and I. E. Farber "The Relation of Anxiety to Differential Eyelid Conditioning," Journal of Experimental Psychology, 47 (1954) 127-134.

--------, I. E. Farber and E. Taylor "The Relation of Electric Shock and Anxiety to Level of Performance in Eyelid Conditioning," Journal of Experimental Psychology, 48 (1954) 404-408.

----------and J. A. Taylor "Anxiety and Strength of the UCS as Determiners of the Amount of Eyelid Conditioning," Journal of Experimental Psychology, 42 (1951) 183-188.

----------and J. A. Taylor "The Relation of Conditioned Response Strength to Anxiety in Normal, Neurotic and Psychotic Subjects," Journal of Experimental Psychology, 45 (1953) 265-272.

----------, J. A. Taylor and R. Ketchel "Anxiety (drive) level and degree of Competition in Paired Associate Learning," Journal of Experimental Psychology, 52 (1956) 306-310.

Stein, H. L. "Needs and Dimensions of Research," Phi Delta Kappan, 37 (1956) 316-317.

Stern, G. G., M. I. Stein and B. S. Bloom.. Methods in Personality Assessment, Glencoe, Ill.; The Free Press, 1956.

Stevenson, G. S. and H. Milt "Ten Tips to Reduce Teacher Tension," National Educational Association Journal, 45 (Dec. 1956) 545-547.

Stopol, M. S. "Rorschach Performance in Relation to two Types of Stress," Journal of Consulting Psychology, 18 (1954) 11-15.

----------"The Consistency of Stress Tolerance," Journal of Personality, 23 (Spt. 1954) 13-29.

Sullivan, Harry Stack. Conceptions of Modern Psychiatry. Washington: The William Alanson White Foundation, 1947.

Sweitzer, R. E. "The Superintendent's Role in Improving Instruction," Administrator's Notebook, 61 (April 1958) 1-4.

Symonds, P. M. "What Education has to Learn from Psychology; VI: Emotion and Learning," Teachers College Record, 60:1 (Oct. 1958) 9-22.

Tallarico, R. B. and E. E. Reitman "Anagram-solving Ability as an Index of Anxiety," Journal of Clinical Psychology, 15:2 (April 1959) 181.

Taylor, J. A. "The Relationship of Anxiety to the Conditioned Eyelid Response," Journal of Experimental Psychology, 41 (1951) 81-82.

----------"A Personality Scale of Manifest Anxiety," Journal of Abnormal and Social Psychology, 48 (1953) 285-290.

----------"Drive Theory and Manifest Anxiety," Psychological Bulletin, 53 (1956) 303-320.

----------and K. W. Spence "The Relationship of Anxiety Level to Performance in Serial Learning," Journal of Experimental Psychology, 44 (1952) 61-64.

Thompson, G. G. and C. W. Hunnicutt "The Effect of Repeated Praise or Blame on the Work Achievement of 'Introverts' and 'Extroverts'," Journal of Educational Psychology, 35 (1944) 257-266.

Thurstone, L. L. and T. G. Thurstone. Psychological Examination for College Freshmen. New York: American Council on Education, Educational Testing Service, 1948.

Valenti, J. J. "Measuring Educational Leadership Attitudes," Journal of Applied Psychology, 36 (Feb. 1952) 36-43.

----------and T. E. Smith "Attitudes of Teachers and Administrators," Administrator's Notebook, 3:1 (Sept. 1954) 1-4.

Walker, H. M. and J. Lev. Statistical Inference. New York: Henry Holt, 1953.

Welsh, G. S. "An Anxiety Index and Internalization Ratio for the MMPI," Journal of Consulting Psychology, 16 (Feb. 1952) 65-72.

Whyte, W. The Organization Man. New York: Simon and Schuster, 1956.

Wilensky, H. "A Factorial Study of Behavioral and Psychological Measures of Anxiety," Journal of Consulting Psychology, 21:3 (June 1957) 216.

Windle, G. "The Relationships Among Five MMPI 'anxiety' Indices," Journal of Consulting Psychology, 19:1 (Feb. 1955) 61-63.

Worth, W. H. "Can Principals Rate Teachers?" The C.S.A. Bulletin, 1:3 (January 1962) to appear.

Yerkes, R. M. and J. D. Dodson "The Relation of Strength of Stimulus to Rapidity of Habit Formation" Journal of Comparative Neurology and Psychology 18 (1908) 459-482 (reported in Eysenck, 1957, pp. 94 ff).

Young, P. T. Motivation and Behavior. New York: Wiley and Sons, 1936.

APPENDIX

TESTS USED IN THE INVESTIGATION

1. The Maudsley Personality Inventory. Neuroticism items: 2, 3, 6, 7, 10, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47. Extraversion items: the remainder.

2. Students' Biographical Index. This is a composite of several scales embedded in non-scored or filler items. Taylor Anxiety Scale items: 5, 9, 16, 26, 31, 35, 40, 48, 51, 56, 58, 61, 64, 68, 71, 75, 77, 84, 88, 93, 95, 97, 105, 110, 117, 120, 125. California Psychological Inventory *Ac* items: 1, 6, 8, 10, 11, 14, 18, 21, 36, 37, 41, 45, 50, 52, 55, 57, 60, 63, 65, 70, 76, 80, 83, 86, 87, 92, 94, 96, 98, 104, 106, 111, 116, 118, 121, 124, 126, 128. California Psychological Inventory *Ai* items: 2, 4, 6, 10, 17, 23, 25, 27, 30, 32, 34, 37, 39, 44, 47, 49, 50, 59, 67, 72, 74, 78, 85, 89, 101, 105, 109, 114, 116, 119. Edwards Personal Preference Schedule need for Achievement items: all those numbered 129 to 156.

3. Confidential Report on Student Teaching (Research Adaptation of Forms 60-2, 50-2), commonly known as the White Rating Sheet.

STUDENTS' BIOGRAPHICAL INDEX

Following is a large number of self-descriptive statements. Here is how to use them:

1. Read each one as though it were describing you, yourself.
2. Decide whether you would agree or disagree with the statement as representing your usual way of acting or feeling.
3. Then, write A (for agree) or D (for disagree) on the blank to the left.
4. Work quickly and don't spend too much time on any one item. Answer each one.
5. If neither A or D seems appropriate, put down whichever one comes closest to describing you, or the way you think or feel.
6. If you find it absolutely impossible to decide, use a question mark (?) but do not use this answer except very occasionally.

NOTE: There are no right or wrong answers; this is not a test. Papers usually contain quite a mixture of A's and D's.

Please make your answers genuine; they will be kept in the strictest confidence, will have no influence whatever on your marks or practice teaching, and will not be related in any official way to your university work here.

1.____I have a very strong desire to be a success in the world.

2____I looked up to my father as an ideal man.

3____I sometimes feel happy, sometimes depressed, without any apparent reason.

4____Our thinking would be a lot better off if we would just forget about words like "probably", approximately", and "perhaps".

5____I am the kind of person who takes things hard.

6____I liked "Alice in Wonderland" by Lewis Carroll.

7____My mind often wanders when I am trying to concentrate.

8____I usually go to the movies more than once a week.

9____I am a very nervous person.

10____I have had very peculiar and strange experiences.

11____I am often said to be hot-headed.

12____I often wish people would be more definite about things.

13____I am frequently "lost in thought" even when I am supposed to be taking part in a conversation.

14____When I was going to school I played hookey quite often.

15____I set a high standard for myself and feel that others should do the same.

16____Life is often a strain for me.

17.___For most questions there is just one right answer, once a person is able to get all the facts.

18.___I think I would like the work of a school teacher.

19___I am sometimes bubbling over with energy and sometimes very sluggish.

20___I find that a well-ordered mode of life with regular hours is congenial to my temperament.

21.___When someone does me wrong I feel I should pay him back if I can, just for the principle of the thing.

22___I am inclined to be moody.

23___I seem to be about as capable and smart as most others around me.

24___It is hard for me to sympathize with someone who is always doubting and unsure about things.

25___I usually take an active part in the entertainment at parties.

26.___I am more self-conscious than most people.

27___The trouble with many people is that they don't take things seriously enough.

28___I often start things I never finish.

29___I have frequent ups and downs in mood, either with or without apparent cause.

30.___It is always a good thing to be frank.

31___At times I feel that I am going to crack up.

32___It is annoying to listen to a lecturer who cannot seem to make up his mind as to what he really believes.

33___My daydreams are frequently about things that can never come true.

34___I don't blame anyone for trying to grab all he can get in this world.

35___I am very confident of myself.

36___Planning one's activities in advance is very likely to take most of the fun out of life.

37___I was a slow learner in school.

38___I am inclined to ponder over my past.

39___I like poetry.

40___I am often sick to my stomach.

41___There is something wrong with a person who can't take orders without getting angry or resentful.

42____I sometimes feel "just miserable" for no good reason at all.

43____I never make judgments about people until I am sure of the facts.

44____Sometimes without any reason or even when things are going wrong, I feel excitedly happy, "on top of the world".

45____I wake up fresh and rested most mornings.

46____I often find that I have made up my mind too late.

47____It is all right to get around the law if you don't actually break it.

48____I am about as nervous as other people.

49____Parents are much too easy on their children nowadays.

50____I have a tendency to give up easily when I meet dificult problems.

51____I work under a great deal of strain.

52____I have the wanderlust and am never happy unless I am roaming or travelling about.

53____A strong person will be able to make up his mind even on the most difficult questions.

54____I have often lost sleep over my worries.

55____I am sometimes cross and grouchy without any good reason.

56____I blush as often as others.

57____My parents have often disapproved of my friends.

58____I have diarrhea ("the trots") once a month or more.

59____Teachers often expect too much work from the students.

60____My way of doing things is apt to be misunderstood by others.

61____I worry quite a bit over possible troubles.

62____I like to have a place for everything and everything in its place.

63____I have had blank spells in which my activities were interrupted and I did not know what was going on around me.

64____When embarrassed I often break out in a sweat which is very annoying.

65____I like to keep people guessing what I'm going to do next.

66____I am often troubled about feelings of guilt.

67____I think I would like to fight in a boxing match sometime.

68____My feelings are rather easily hurt; that is, easier than most people.

69____I don't like to work on a problem unless there is the possibility of coming out with a clear-cut and unambiguous answer.

70____If given the chance I would make a good leader of people.

71____I do not often notice my heart pounding and I am seldom short of breath.

72____I like to plan a home-study schedule and then follow it.

73____I would tend to rate myself as a tense or "highly-strung" individual.

74____I have often found people jealous of my good ideas, just because they had not thought of them first.

75____I have periods of such great restlessness that I cannot sit long in a chair.

76____In school I was sometimes sent to the principal for cutting up.

77____I feel anxious about something or someone almost all of the time.

78____People pretend to care more about one another than they really do.

79____I often experience periods of lonliness.

80____I like to read about history.

81____It bothers me when something unexpected interrupts my daily routine.

82____I like to indulge in reverie (daydreaming).

83____I am so touchy on some topics that I can't talk about them.

84____Often my bowels don't move for several days at a time.

85____The future is too uncertain for a person to make serious plans

86____I don't seem to care what happens to me.

87____I like to talk before groups of people.

88____My sleep is restless and disturbed.

89____The man who provides temptation by leaving valuable property unprotected is about as much to blame for its theft as the one who steals it.

90____Most of the arguments or quarrels I get into are over matters of principle.

91____I spend much time in thinking over good times I have had in the past.

92____I am often bothered by useless thoughts which keep running through my mind.

93____I often dream about things I don't like to tell other people.

94____I like to plan out my activities in advance.

95____I often find myself worrying about something.

96____I must admit I find it very hard to work under strict rules and regulations.

97____I wish I could be as happy as others.

98____I like large, noisy parties.

99____I am known as a hard and steady worker.

100____I have often felt listless and tired forno good reason.

101____I sometimes feel that I am a burden to others.

102____I don't like things to be uncertain and unpredictable.

103____After a critical moment is over, I often think of something I should have done but failed to do at the time.

104____I certainly feel useless at times.

105____I do not have as many fears as my friends.

106____I always try to do at least a little better than what is expected of me.

107____Once I have made up my mind I seldom change it.

108____Ideas run through my head so that I sometimes find it hard to sleep.

109____Lawbreakers are mostly always caught and punished.

110____I have often felt that I faced so many difficulties that I could not overcome them.

111____I like to plan out my activities in advance.

112____I think I am stricter about right and wrong than most people.

113____I have sometimes been bothered by having a useless thought come into my mind repeatedly.

114____I dread the thought of an earthquake.

115____I am in favour of a very strict enforcement of all laws, no matter what the consequences.

116____I often lose my temper.

117____At times I have been worried beyond reason about something that really did not matter.

118____My parents were always very strict and stern with me.

119____I am bothered by people outside, in busses, in stores,etc., watching me.

120____I don't like to face a difficulty or make an important decision.

121____I often get disgusted with myself.

122____I always see to it that my work is carefully planned and organized.

123____I often feel disgruntled.

124____Society owes a lot more to the businessman and manufacturer than it does to the artist and the professor.

125____I am not at all confident of myself.

126____I think I would like to belong to a motorcycle club.

127____People who seem unsure and uncertain about things make me feel uncomfortable.

128____I used to like it very much when one of my papers was read to the class at school.

- -

STOP READ THESE NEW DIRECTIONS BEFORE COMPLETING THE NEXT 28 ITEMS.

Following are 28 pairs of statements about things that you may or may not like, or about ways in which you may or may not feel. Decide which statement in each pair is more characteristic of what you like or how you feel. Put a circle around the letter (A or B) in front of that statement. If you agree with both statements, choose the one that is more characteristic. If you think that neither accuractely describes what you like or how you feel, choose the one you consider to be less inacurate.Don't skip any.

129 A. I like to solve puzzles and problems that other people have difficulty with.
B. I like to follow instructions and to do what is expected of me.

130 A. I like to find out what great men have thought about various problems in which I am interested.
B. I would like to accomplish something of great significance.

131 A. I would like to write a great novel or play.
B. I like to attack points of view that are contrary to mine.

132 A. I like to read newspaper accounts of murders and other forms of violence.
B. I would like to write a great novel or play.

133 A. I like to be able to do things better than other people can.
B. I like to tell amusing stories and jokes at parties.

134 A. Any written work that I do I like to have precise, neat and well organized.
B. I would like to be a recognized authority in some job, profession, or field of specialization.

135 A. I like to be able to say that I have done a difficult job well.
B. I like to work hard at any job I undertake.

136 A. I like to be one of the leaders in the organizations and groups to which I belong.
B. I like to be able to do things better than other people can.

137 A. I like to be loyal to my friends.
B. I like to do my very best in whatever I undertake.

138 A. I would like to accomplish something of great significance.
B. I like to attract persons of the opposite sex.

139 A. I would like to write a great novel of play.
B. When serving on a committee, I like to be appointed or elected chairman.

140 A. I like to work hard at any job I undertake.
B. I would like to accomplish something of great significance.

141 A. I like to do my very best in whatever I undertake.
B. I like to help other people who are less fortunate than I am.

142 A. I like to observe how another individual feels in a given situation.
B. I like to be able to say that I have done a difficult job well.

143 A. I would like to be a recognized authority in some job, profession or field of specialization.
B. I like to have my work organized and planned before beginning it.

144. A. I like to tell amusing stories and jokes at parties.
B. I would like to write a great novel or play.

145. A. I like to be successful in things undertaken.
B. I like to form new friendships.

146. A. I like to solve puzzles and problems that other people have difficulty with.
B. I like to judge people by why they do something -- not by what they actually do.

147 A. I like to accomplish tasks that others recognize as requiring skill and effort.
B. I like my friends to encourage me when I meet with failure.

148 A. I like to be able to come and go as I want.
B. I like to be able to say that I have done a difficult job well.

149 A. I like to be able to do things better than other people can.
B. I like to eat in new and strange restaurants.

150 A. I like to travel and to see the country.
B. I like to accomplish tasks that others recognize as requiring skill and effort.

151. A. I would like to be a recognized authority in some job, profession or field of specialization.
B. I feel guilty whenever I have done something I know is wrong.

152 A. When things go wrong for me, I feel that I am more to blame than anyone else.
B. I like to solve puzzles and problems that other people have difficulty with.

153 A. I like to accomplish tasks that others recognize as requiring skill and effort.
B. I like to be able to come and go as I want.

154 A. I like my friends to encourage me when I meet with failure.
B. I like to be successful in things undertaken.

155 A. I like to help my friends when they are in trouble.
B. I like to do my very best in whatever I undertake.

156 A. I like to go out with attractive persons of the opposite sex.
B. I like to be successful in things undertaken.

- - - - - - - - -

MAUDSLEY PERSONALITY INVENTORY

(Copyright 1959: H.J. Eysenck)

Name____

Age______ Sex_____ Occupation__________________________

N: E: ?:

Instructions:

Here are some questions regarding the way you behave, feel and act. After each question there is a "Yes", a "?" and a "No".

Try to decide whether "Yes" or "No" represents your usual way of acting or feeling; then put a circle around the "Yes" or "No". If you find it absolutely impossible to decide, put a circle around the "?", but do not use this answer except very occasionally. Work quickly and don't spend too much time over any question; we want your first reaction, not a long drawn-out thought process ! The whole questionnaire shouldn't take more than a few minutes. Be sure not to omit any questions. Now go ahead, work quickly, and remember to answer every question. There are no right or wrong answers, and this isn't a test of intelligence or ability, but simply a measure of the way you behqve.

1. Are you happiest when you get involved in some project that calls for rapid action ? Yes ? No
2. Do you sometimes feel happy, sometimes depressed, without any apparent reason ? Yes ? No
3. Does your mind often wander when you are trying to concentrate ? Yes ? No
4. Do you usually take the initiative in making new friends? Yes ? No
5. Are you inclined to be quick and sure in your actions ? Yes ? No
6. Are you frequently "lost in thought" even when supposed to be taking part in a conversation ? Yes ? No
7. Are you sometimes bubbling over with energy and sometimes very sluggish ? Yes ? No
8. Would you rate yourself as a lively individual ? Yes ? No
9. Would you be very unhappy if you were prevented from making numerous social contacts ? Yes ? No
10. Are you inclined to be moody ? Yes ? No
11. Do you have frequent ups and downs in mood, either with or without apparent cause ? Yes ? No

12. Do you prefer action to planning for action ? Yes ? No

13. Are your daydreams frequently about things that can never come true ? Yes ? No

14. Are you inclined to keep in the background on social occasions ? Yes ? No

15. Are you inclined to ponder over the past ? Yes ? No

16. Is it difficult to "lose yourself" even at a lively party ? Yes ? No

17. Do you ever feel "just miserable" for no good reason at all ? Yes ? No

18. Are you inclined to be overconscientious ? Yes ? No

19. Do you often find that you have made up your mind too late ? Yes ? No

20. Do you like to mix socially with people ? Yes ? No

21. Have you often lost sleep over your worries ? Yes ? No

22. Are you inclined to limit your acquaintances to a select few ? Yes ? No

23. Are you often troubled about feelings of guilt ? Yes ? No

24. Do you ever take your work as if it were a matter of life or death ? Yes ? No

25. Are your feelings rather easily hurt ? Yes ? No

26. Do you like to have many social engagements ? Yes ? No

27. Would you rate yourself as a tense or "highly-strung" individual ? Yes ? No

28. Do you generally prefer to take the lead in group activities ? Yes ? No

29. Do you often experience periods of lonliness ? Yes ? No

30. Are you inclined to be shy in the presence of the opposite sex ? Yes ? No

31. Do you like to indulge in reverie (daydreaming) ? Yes ? No

32. Do you nearly always have a "ready answer" for remarks directed at you ? Yes ? No

33. Do you spend much time in thinking over good times you have had in the past ? Yes ? No

34. Would you rate yourself as a "happy-go-lucky" individual? Yes ? No

35. Have you often felt listless and tired for no good reason? Yes ? No

36. Are you inclined to keep quiet when out in a social group ? Yes ? No

37. After a critical moment is over, do you usually think of something you should have done but failed to do ? Yes ? No

38. Can you usually let yourself go and have a hilariously good time at a gay party ? Yes ? No

39. Do ideas run through your head so that you cannot sleep ? Yes ? No

40. Do you like work that requires considerable attention ? Yes ? No

41. Have you ever been bothered by having a useless thought come into your mind repeatedly ? Yes ? No

42. Are you inclined to take your work casually, that is as a matter of course ? Yes ? No

43. Are you touchy on various subjects ? Yes ? No

44. Do other people regard you as a lively individual ? Yes ? No

45. Do you often feel disgruntled ? Yes ? No

46. Would you rate yourself as a talkative individual ? Yes ? No

47. Do you have periods of such great restlessness that you cannot sit for long in a chair ? Yes ? No

48. Do you like to play pranks upon others ? Yes ? No

University of Alberta, Faculty of Education

CONFIDENTIAL REPORT ON STUDENT TEACHING

(Research adaptation of Forms 60-2, 50-2)

Student Teacher________________________ Acting Faculty Consultant__________

GENERAL EVALUATION		-3	-2	-1	0	1	2	3	Difference
(impression of lesson as a whole; not to be derived by summing ratings of the components)	Lesson One				✓				
	Lesson Two								

- -

PERSONAL QUALITIES		-3	-2	-1	0	1	2	3	Difference
(poise, voice, manner, professional interest, attitude)	Lesson One				✓				
	Lesson Two								

PREPARATION FOR TEACHING		-3	-2	-1	0	1	2	3	Difference
(academic knowledge, suitability of lesson materials, planning)	Lesson One				✓				
	Lesson Two								

PRESENTATION OF LESSONS		-3	-2	-1	0	1	2	3	Difference
(effectiveness of methods, adaptability, command of English, resourcefulness)	Lesson One				✓				
	Lesson Two								

CONTACT WITH PUPILS		-3	-2	-1	0	1	2	3	Difference
(sensitivity to pupil reaction, adjustment to grade level, manner of dealing with difficulties)	Lesson One				✓				
	Lesson Two								

CLASSROOM MANAGEMENT		-3	-2	-1	0	1	2	3	Difference
(control of class, giving directions, handling routine matters, keeping records)	Lesson One				✓				
	Lesson Two								

B29796